NATIONAL QUALITY FORUM

+ + + + +

RENAL ENDORSEMENT MAINTENANCE
STEERING COMMITTEE

+ + + + +

TUESDAY
AUGUST 16, 2011

+ + + + +

The Steering Committee met at the Marriott Metro Center, 775 12th Street, N.W., Washington, D.C., at 9:00 a.m., Peter Crooks, Co-Chair, presiding.
PRESENT:

PETER CROOKS, MD, Co-Chair
KRISTINE SCHONDER, PharmD, Co-Chair *
CONSTANCE ANDERSON, BSN, MBA, Northwest Kidney Centers
JEFFREY BERNS, MD, University of Pennsylvania School of Medicine
LORIEN DALRYMPLE, MD, MPH, University of California Davis Medical Center *
ANDREW FENVES, MD, Baylor Health Care System
MICHAEL FISCHER, MD, MSPH, Department of Veterans Affairs, University of Illinois
JERRY JACKSON, MD, Nephrology Associates, PC
FREDERICK KASKEL, MD, PhD, Children's Hospital at Montefiore
MYRA KLEINPETER, MD, MPH, Tulane University School of Medicine
ALAN KLIGER, MD, Hospital of St. Raphael/Yale University School of Medicine
LISA LATTS, MD, MSPH, MBA, WellPoint, Inc.
KATHE LeBEAU, Renal Support Network
STEPHEN D. MCMURRAY, MD, DaVita, Inc.
JOSEPH V. NALLY, JR., MD, Cleveland Clinic Foundation

ANDREW NARVA, MD, National Institute of

Diabetes and Digestive and Kidney Diseases, National Institutes of Health
JESSIE PAVLINAC, MS, RD, CSR, LD, Oregon Health and Science University
MICHAEL SOMERS, MD, Children's Hospital Boston
RUBEN VELEZ, MD, Dallas Nephrology Associates
ROBERTA WAGER, RN, MSN, American Association of Kidney Patients
JANET WELCH, PhD, RN, Indiana University School of Nursing
HARVEY WELLS, Dialysis Patient Advocate

NQF STAFF:

HELEN BURSTIN, MD, MPH
TENEE DAVENPORT
KAREN PACE, PhD, RN
LAUREN RICHIE, MA

ALSO PRESENT:

ASHFAQ AKHTAR, Amgen
MUREEN ALLEN, ActiveHealth Management *
KATHERINE AST, American Medical Association
KERI CHRISTENSEN, American Medical Association
BARBARA FIVUSH, American Society of Pediatric Nephrology
EDWARD JONES, Renal Physicians Association
DIEDRA JOSEPH, American Medical Association
LISA MCGONIGAL, Kidney Care Partners
JOSEPH MESSANA, CMS
KATHRYN SCHUBERT, American Society of Pediatric Nephrology
ROBERT WOLFE, CMS

*Participating via teleconference

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

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

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

9:03 a.m.

CO-CHAIR CROOKS: Welcome, everybody.

I am Peter Crooks. It has been my honor to chair this process now a couple of times, a third time. You know, the third time is a charm. I think, hopefully, this time we will get it right, because it is always challenging and fun.

But, on behalf of myself and my Co-Chair Kristine Schonder, who can't be here today -- she will be calling in this morning -- welcome. And thank you all for being here and for participating in this important work.

I am going to try to keep my comments short, so we can get busy.

I just want to say a couple of things that I think we have all been talking about this morning, the new and improved NQF evaluation process for Steering Committees. I think what it is really saying to me is the

process is evolving. Those of us who were here three years ago, and then in January, and now once again, we have been able to see the process is tightening up, and more is expected from measure developers, and more in the way, of course, of validity and reliability. I think in the long run that is a good thing.

But, as I told Karen, why is it that this Committee is always the pilot case? I don't know. That has been our good fortune before.

But the importance has been better defined. Impact, does it have high impact or not? That is really very important. Is there a gap in care? What does the evidence say? Does it really support it or not? And some guidance, so I think it is very helpful on how to rate the evidence and whether it supports the metric.

Then, what I think is the biggest challenge is coming to grips with reliability and, more difficult, validity. I am sure we

will be talking about that to some extent as we try to apply these new criteria. Now it is actually a stop criteria. If it is not valid, we can't accept it.

So, that is going to be the challenge, I think, for this group today. And hopefully, we can help the NQF figure out how to keep improving the process for this.

So, a few announcements. First of all, remember the meeting is open to the public and it is being audiotaped. Please use your microphones.

I don't think the request function works on these. So, you just have to raise your hand, and Karen and Lauren will help me see the field. If I am involved over here, I might not see you over here, but we will try to scan the deck.

What else? The schedule I think is known to all of you. Today we are starting at 9:00. Tomorrow we will start an hour earlier. Breakfast will start at 7:30. We are going

to try to be done by 3:00 or 3:15, so we can catch those 5:30 flights from Dulles, well, speaking personally.

We have several new Committee Members. Lorien Dalrymple is going to be calling in. I guess because of pregnancy, she is unable -- she has had a baby. So, she will be able to call and participate a little later.

DR. DALRYMPLE: Hi.

CO-CHAIR CROOKS: Oh, you're on?

DR. DALRYMPLE: Hi.

CO-CHAIR CROOKS: Hi, Lorien.

DR. DALRYMPLE: Good morning.

CO-CHAIR CROOKS: So, we will be asking you to introduce yourself a little bit later, along with Dr. Andrew Fenves, Michael Fischer, Stephen McMurray, Michael Somers, and Janet Welch.

And thank you all for volunteering to participate. You didn't know what you were getting into, but you will find out pretty

fast.

(Laughter.)

Kristine, are you on?

(No response.)

She will be calling in later? Okay.

I would also like to welcome the measure stewards and developers who are with us today. I won't go through the list. In the past, you may recall we have had two or three developers of our metrics. Now we are up to seven. And they will all be introducing themselves a bit in turn as their metrics come up.

Okay. So, I think at this point I can turn it over to Helen for her greeting.

Are you ready to greet?

DR. BURSTIN: Sure.

CO-CHAIR CROOKS: Okay. Thank you.

DR. BURSTIN: Good morning, everybody. Welcome again. Welcome back to more renal.

We again want to thank you for coming back and participating one more time and, again, want to thank you as well for helping us as we move through our new processes. We hope they add clarity, but I think, as we have been learning through the first several Steering Committees who have used the new criteria, you will be, I think, relying fairly heavily on Karen to help interpret some of that.

We did put together a small cheatsheet -- have we passed it out yet? -- okay, that we will pass out that just very simply reviews each of the criteria and the ratings that will be associated when you are asked to vote, just to kind of keep it very simple.

Please let us know if this is useful. We literally just put it together. We just thought it might be nice to just have a very simple -- Karen and I initially kept talking about the imaginary one-pager. It

doesn't go on one page, but it is two-sided, you know, two back-to-back, two pages. So, we hope this is useful.

It also explains exactly where the stop sequences are now. So, the last time you met, importance to measure and report was a must-pass criterion, and if it didn't pass that, particularly about gaps, the gap, or especially any issues around evidence, we stopped evaluating the measure.

An important new feature of the work for the Task Forces, and ultimately passed by the Board as well, is that now we also have a stop after scientific acceptability. So, if the measure is not reliable and valid with precise specifications, it doesn't really matter if it is usable or feasible, either. So, we have added that to the hierarchy.

So, I think it will be a slightly different process today. We can distribute these.

The only other thing I will mention is, also, our General Counsel, Ann Hammersmith, is on vacation this week. So, I am happy to give you the brief intro to disclosures.

So, as you are doing the introductions this morning, please disclose anything you think is important for your fellow Committee members to know about. We don't need you to recite your CVs. We all have read them. They are voluminous. But really indicate areas particularly where you think there might be conflicts with any of the measures. And at the end, we will ask you if you have any questions for each other as we go through this process.

So, with that, I will turn it back to Peter and we can go around the table.

CO-CHAIR CROOKS: Okay. Thank you, Helen.

I think we are ready to do introductions then. Okay. So, I will start

out with an exemplary introduction, I hope.

(Laughter.)

Behind us is a few points we would like you to touch on today, and just one non-medical interest to sort of broaden out our view of you.

So, I will start out. My name is Peter Crooks. I am with Kaiser Permanente in Southern California. I live in Los Angeles and have survived an earthquake and a lot of other natural disasters over the years.

I have been involved in quality since the early nineties when we began to develop our quality program at Kaiser Permanente and have been fortunate enough to be involved with the KDOQI committees, Steering Committee, several companies that were evolved with Kaiser and Fresenius to bring quality to the marketplace, and now with the National Quality Forum.

So, my main non-medical interest is music, as a composer and a performer. I

didn't bring any tapes today, I'm sorry to say, but I will next time, if there is a next time.

And on to my disclosures, under have I had any direct relationship with any organization of the types listed in the disclosure-of-interest policy, I am on the Board of Directors of the California Dialysis Council, which is a political action committee informing the legislature about the needs of dialysis patients and the dialysis industry.

As a partner at Kaiser Permanente, I do help develop quality metrics, but nothing has been submitted this go-round and nothing really directly competing.

My partners serve as medical directors in numerous Fresenius medical care facilities as well as one DaVita facility and one Renal Advantage facility.

So, I think that is about it for me.

Okay. Shall we just move around to

the right? Okay, Janet?

DR. WELCH: My name is Janet Welch. I am a Professor of Nursing at Indiana University School of Nursing in Indianapolis.

My area of research interest is self-management of diet and fluid limitations by hemodialysis patients.

And I would say a non-medical interest is probably crafting.

The only disclosure I had listed was that I am the Chair of the Membership Committee for the Midwest Nursing Research Society.

DR. PAVLINAC: Good morning.

Jessie Pavlinac from Portland, Oregon. I'm the Director of Inpatient Clinical Nutrition at Oregon Health and Sciences University, Hospitals and Clinics.

My quality interest and experience, I was Chair of the Chronic Kidney Disease Evidence Analysis Process for the American Dietetic Association a couple of years and

have worked in other quality areas.

My non-medical interest, I am a charter member of the One More Time Around Again Marching Band -- (laughter) -- since Peter started out with a music gig, and an alumni marcher for Oregon State University.

Disclosure: I am the current Vice Chair for the Northwest Renal Medical Review Board up in Seattle.

How are you, Connie?

And I have been a member of the Oregon and National Kidney Foundation Council on Renal Nutrition.

DR. SOMERS: I'm Michael Somers. I am a pediatric nephrologist from Children's Hospital in Boston, where I am the Director of Clinical Service and help direct the dialysis unit as well.

Several years ago, I got a phone call from someone in the Hospital asking me if I would like to be involved with something called quality, and that is how I first began

to have an interest in this realm. I have been involved at Children's helping to formulate quality plans for our Division and portions of the Hospital as a whole.

In terms of my non-medical interests, I like to kayak. I also like to do trivia. And actually, I like to kayak doing trivia with a couple of people.

(Laughter.)

So, I know that is more than one, but it combines things together.

In terms of disclosures, I am a member of an Advisory Board for Novartis, and I am also the Treasurer and am on the Board of Directors for the ESRD Network in New England.

DR. FISCHER: My name is Michael Fischer. I am a clinical nephrologist at the University of Illinois and the Department of Veterans Affairs in Chicago.

I became involved in this because the VA had asked me to come and kind of be the VA representative, as we have recently formed

a Chronic Kidney Disease Working Group looking at developing performance measures within the Department of Veterans Affairs and capitalizing on the VA electronic information system.

My non-medical interest, a big tennis fan. I hope to go to all four of the Grand Slams one day. Two down, two to go.

I think my conflicts of interest, I have the customary society memberships. And other than the VA Working Group, which we just kind of started, I don't think it is in any competition with this organization.

DR. LATTS: Good morning.

I'm Lisa Latts. I'm Vice President for Public Health Policy with WellPoint, which is a large national health insurer.

I am an internist, have a subspecialty in medical complications of pregnancy. And I'll tell you, when you assigned me the measures that involve the parathyroid calcium phosphorus access, I was

like, oh, my God. Thanks.

(Laughter.)

I have been in charge of quality for some time with WellPoint, although now am more on the public health sector working on our public health programs.

I live in Denver, Colorado. Non-work interest, my hobby is travel, ideally, overseas and as exotic as possible. Although having 21-month-old twins means that I don't get to do it very much.

Oh, other thing I should add for quality interest is that I am also a renal patient, having developed HUS after I delivered my twins, on ESRD and dialysis for about 10 months, and now I am about 10 months post-kidney-transplant.

Then, the only other disclosure is that I work for WellPoint, obviously.

DR. NALLY: Good morning.

I'm Joe Nally. I am the Director of the Center for Chronic Kidney Disease at

the Cleveland Clinic, which makes me a Clevelander, some good, some bad.

My interest in quality relates back to the original KDOQI days, where I have been part of that effort since 2002. I am currently the Vice Chair of Public Policy, KDOQI, for NKF.

In terms of other interests and disclosures, I am the PI at the Clinic for our CKD Registry of over 60,000 CKD patients. That was originally started with an unrestricted grant from Amgen and now Genzyme.

Non-medical interest, what can you say after Lisa's non-medical interest? I am simply a golfer, a racquetball player, and support the travel of my children.

(Laughter.)

DR. JACKSON: Good morning.

I am Jerry Jackson from Birmingham, Alabama. I am a practicing nephrologist there.

I am involved a great deal with

interventional nephrology and, also, am the Medical Director of two Fresenius dialysis clinics.

I have been involved with Network 8 for over 15 years and am Chairman of the Medical Review Board there. Got involved in quality, management quality interests, largely through that. I have been on the Network Forum. Served on the Quality, Safety, and Accountability Subcommittee of the Renal Physicians Association.

As far as outside interests, primarily right now grandchildren -- we have four -- and, also, photography and gardening.

DR. KASKEL: Hi. I'm Rick Kaskel, pediatric nephrologist at the Children's Hospital at Montefiore of Albert Einstein; Vice Chair and Director of Pediatric Nephrology and Child Health. Have some research interest in climara disease progression and chronic kidney disease in children.

The only disclosure is some NIH support.

And outside interests are my family and I like to sail.

DR. BERNS: Good morning.

Jeff Berns from the University of Pennsylvania in Philadelphia.

I am here as a representative from the American Society of Nephrology. I have been involved, as Joe, in KDOQI, actually, when it was DOQI, then KDOQI. And I am Vice Chair for Practice Guidelines and Commentaries for the NKF.

Disclosures: I have been involved as an advisor for clinical trials for Amgen, which is not an active endeavor. And I think that is the only disclosure at this time.

Non-medical interests, I am training for the New York Marathon.

DR. FENVES: Good morning.

My name is Andrew Fenves. I'm from Dallas. I'm an adult nephrologist. I

represent Baylor Healthcare Systems. I am Chief of the Division of Nephrology there.

I got involved because my institution is doing a lot of new safety and quality review, and they wanted me to get involved.

Obviously, I'm from Dallas. My outside interest, according to my wife, I'm addicted to duplicate bridge, which is true, but I only get to play online and occasionally in person.

And my disclosures: I have grant support from the Baylor Cancer Center and the NIH, and I am Co-Editor for a few more months of Dialysis and Transplantation, which, unfortunately, is closing in October.

DR. KLIGER: I'm Alan Kliger. I am a nephrologist in New Haven, Connecticut. I'm the Chief Medical Officer and Chief Quality Officer for the Hospital of St. Raphael in New Haven.

My quality interest really started,

I was on the Steering Committee of the original DOQI, which then became KDOQI, and have been involved with quality since then. Have served as the Chair of the RPA's Quality, Safety, and Accountability Committee.

Non-medical interest, I sing. I sang with the New Haven Chorale in Europe this summer, which was wonderful.

And in terms of disclosures, I also have some support from the NIH and some support for investigator-directed research from Amgen. And I am a member of the Board of Directors of the Renal Physicians Association.

DR. VELEZ: Ruben Velez. I'm a nephrologist in Dallas. I have been there for 30-something years, coming originally from Puerto Rico. So, you can feel my Texan accent here.

(Laughter.)

I'm still practicing. One day I'll get it.

But, anyway, Medical Director of

Fresenius facilities, clinical mostly. My main disclosure would be President of the Renal Physicians Association. Really no other conflict of interest at this time.

Outside medicine, I definitely love sailing and scuba diving, which I was able to do with my family after trying to get them together. So, I am trying to spend more time with the family, and that is a project by itself.

DR. MCMURRAY: Hi. I'm Steve McMurray. I'm a nephrologist, live in Scottsdale, Arizona. I am VP of Clinical Integrated Care Management Services for DaVita.

My interest quality, I was on the Review Board of Network 9 and 10 for about 20 years and served there, and was on the Renal Physicians Association and helped work on some of the quality measures during that time.

My interests, I like to play golf; I like to collect contemporary art. Those are

the two things that keep me going during the rest of the days.

I think my only disclosure is that I do work for DaVita.

DR. KLEINPETER: Hi. I'm Myra Kleinpeter. I am a nephrologist from Tulane in New Orleans.

I got involved in quality originally as Director of the Outpatient Clinics at Charity Hospital in New Orleans before Katrina and did a lot of the projects related to ambulatory care. And since things changed, we now do primarily nephrology and have been involved in the Network 13 Quality Improvement Committee.

My disclosures: I'm on the Speakers' Bureau for Pfizer, Gilead, Glaxo, Boehringer, and some things coming up soon with Amgen.

In terms of my non-medical interest, I like to travel, but this summer it has been hot everywhere I have gone.

(Laughter.)

So, we'll try the winter this year and see if we can get a little bit better travel things done.

And that's it. Thank you.

MS. ANDERSON: I'm Connie Anderson from the Northwest Kidney Centers in Seattle. I am responsible, as Vice President of Clinical Operations, for the quality programs at the Northwest Kidney Centers. I also staff the Quality Committee of the Board of Trustees that oversees all of our quality programs. So, I have been embedded and passionate about quality for many, many years.

I also serve on the Quality Committee of the National Renal Administrators Association and with KCP, the Kidney Care Partners.

In terms of my non-medical interest, well, my passion is snow skiing, but just recently I had the opportunity to perform on stage in Guys and Dolls. So, I think I may

change my passion. It was a great experience.

In terms of disclosures, I don't think I have any.

MS. WAGER: My name is Bobbie Wager. I'm a nephrology nurse/treatment options specialist in San Antonio, Texas.

My disclosure is I work with Fresenius Medical Care. I have been a patient advocate for about 30-some years, since my first transplant. I am a two-time transplant recipient and was hemodialysis.

Non-medical interest, my husband and son and I have four beautiful Scottish terriers. So, does it matter which order? Sometimes the dogs come first.

(Laughter.)

I'm sorry, they do.

I'm an avid Illinois fan, football, go Illini, and avid Washington Redskins fan. So, I love sports.

MS. LEBEAU: Hi. I'm Kathie LeBeau. I am the Patient Advocacy Project

Manager for the Renal Support Network, a national patient group run by patients.

My interest in quality, frankly, is self-interest and that of my fellow patients because I am a home hemodialysis patient the past four years, three years now, and a waiting transplant candidate.

Did I mention I'm from Albany, New York? Yes.

My non-medical interest, well, most of you who were here in January know that, although I am a very serious patient advocate, I am a professional clown part-time. I play symphonic kazoo. So, I am very interested in sharing that with the folks who have musical interest in the room.

My disclosures: I participate in a number of renal coalitions and committees, UNOS and the ESRD Network of New York.

And I think that is everything.

MR. WELLS: My name is Harvey Wells. I have no position, nor am I

organized.

(Laughter.)

I live in between Dallas and Fort Worth.

My interest in quality is I, too, am a patient. I found out I had some renal insufficiency when I was 18 and I tried to join the Navy. Other than that, I had no outward signs. So, I was classified 4F, and I really did nothing about it until my mid-forties when I went to a nephrologist because of high blood pressure. My doctor wanted me to have a biopsy done, and I found out that I was going to be on dialysis within six months. I was able to put it off for four years, changing some practices of mine, and what have you.

But, eventually, I was on dialysis for six months. My wife donated her kidney, and it lasted eight years. She wouldn't give me the other one.

(Laughter.)

I went back on dialysis for five-and-a-half years, and I thought that is how my life was going to end. But, fortunately, since we have been together last, I had a transplant in March at Baylor University in Dallas, and it is working great. My life has changed again.

My interest in quality, I found out then a lot of things about dialysis and kidney care. And over the last four years, I have spent a lot of time traveling to dialysis centers and talking to patients and just getting their perspective. And I have tried to encourage more dialysis patients to consider home options because I feel that they deliver a better quality of care. They are able to help you to live your live like you had originally wanted to, and it certainly did me.

My non-medical interest, I love to travel and I have four grandchildren that I love spending time with. Originally, when I

started traveling for NxStage Medical, I thought it was going to be the latter part of my life. I told somebody, "I just didn't realize it might be the longest chapter of my life." And I spend a lot of time traveling with my grandchildren and visiting different places.

I am a Cleveland Browns fan, a suffering Cleveland Browns fan.

(Laughter.)

And one year, I actually was able to attend all their games, home and away.

The only disclosure I have is I am paid by NxStage when I represent them at the centers.

I appreciate being here.

CO-CHAIR CROOKS: Lorien? Yes, Lorien, please go ahead.

DR. DALRYMPLE: Okay. My name s Lorien Dalrymple. I am a nephrologist at UC Davis. I spend the majority of my time doing clinical research. I am an epidemiologist.

My non-clinical interest I would say is cooking, but my husband would probably disagree. I cook about once every six months now.

As mentioned, I have a 12-day-old son and year-old daughter. I am sorry I couldn't join you in person.

The only disclosure I have is that one of the sources of research funding I receive is from Dialysis Clinics, Incorporated.

CO-CHAIR CROOKS: Okay. Thank you.

Andrew Narva will be joining us, I understand, in a bit. He is coming, but was delayed.

And is Kristine on now?

Are you on, Kristine?

(No response.)

Okay.

MS. RICHIE: Hi, everyone.

I'm Lauren Richie, the Project Manager, now in my second tour of renal duty.

(Laughter.)

I would like to thank everyone for coming again. And to the new Members, thank you. And thank you for putting up with my slew of emails over the last few months.

DR. PACE: And I'm Karen Pace. I'm a Senior Program Director at NQF.

And again, I also would like to thank you for all your hard work and preparation for this meeting.

And we have one other staff person here, Tenee Davenport, who will be helping us with the electronic voting today.

So, I guess, with that, we can get into our program.

Oh, we need to ask if anyone has any questions.

DR. BURSTIN: Just briefly, based on what you have heard, does anybody have any questions of each other about your disclosures? Anything you would like to bring up or raise?

Yes, please.

CO-CHAIR CROOKS: I have one minor disclosure to add that hit me as I was listening to other people. I also have some NIH support on a grant on studying racial disparities in CKD care.

Okay. Any other comments or questions for each other?

(No response.)

Okay. I think it is time, then, for -- and we are doing well; we are on time -- for Karen and Lauren to --

DR. DALRYMPLE: Hello.

CO-CHAIR CROOKS: Hello?

DR. DALRYMPLE: I'm sorry. It's Lorien.

I receive NIH funding, research funding, for the UC Davis Clinical and Translational Science Center. I don't think it is relevant to this. I just wanted to add that.

CO-CHAIR CROOKS: Okay. Thank you.

Any other last-minute disclosures? Come clean now.

(No response.)

Okay. I will turn it over to Karen and Lauren to review our project and kick us off today.

MS. RICHIE: Okay. Karen and I are going to start with an overview of the project. Then, Karen will go into a little bit deeper details as far as the actual measure criteria and evaluation process.

So, most of you were on the orientation call. So, we will keep this introduction very brief.

Just as a reminder, the purpose of this project is to, again, identify and endorse renal-related measures for public reporting and quality improvement, and a little bit different from the previous project in that we are also looking at currently-endorsed measures for maintaining their endorsement status.

We will see, too, this project, different from the last project as well, which was specific to ESRD. This project now includes CKD and ESRD. Although we did include a call for other renal-related conditions, we did not receive any measures. So, again, primarily CKD and ESRD.

Again, the Steering Committee is asked to act as a proxy for our multi-stakeholder membership; work with us here at NQF to achieve the goals of the project; evaluate the submitted measures against our criteria, which we know is a little bit different from the last time, and then to make recommendations to the NQF membership for endorsement, as well as respond to comments received on the measures once they go out for comment for our public and NQF members.

Again, this is just a visual schematic of our consensus-development process, or our CDP, as we like to refer to it. There you can see the project Steering

Committee's role, followed by drafting recommendations, and so forth.

Again, the objectives for today's meeting, we will spend the better part of today and tomorrow evaluating the measures against the new criteria. Then, tomorrow we will get into evaluating measures for related and competing measures, as well as identify gaps in performance measures.

So, just a high-level overview of the measures that we have: 34 in total with the bulk of them being around anemia, cardiovascular, dialysis adequacy, mineral metabolism, and vascular access. We do have one mortality measure and a combination of patient education and quality-of-life measures.

So, with that, I am going to turn it over to Karen for our measure evaluation criteria.

DR. PACE: Okay. I just wanted to review a few things. We talked about some of

this on the orientation call and, then, on our optional call, and had talked with Peter last week and he suggested that I spend a little time trying to get us all on the same page, specifically about reliability and validity. But I will also touch on some of the recommendations from the Task Force about evidence. So, just to try to get us started out on the same page, and then we will work through the individual measures.

So, one of the things that Lauren mentioned is that in this project we will be looking at both new measures and endorsed measures. So, I wanted to just kind of explain that.

In our process now, endorsed measures are required to meet the same criteria, the current criteria that new measures would be. So, even though the endorsed measures were endorsed in 2007 or later, as Peter mentioned, our criteria and guidance, mainly guidance on how we apply the

criteria has evolved over the years. Measures, whether they are endorsed or new, are expected to meet the criteria that are current at that point in time.

There are a few things that we try to focus on maybe a little bit differently with the endorsed measures. The first one is that, hopefully, an endorsed measure has been implemented. And if so, we would like to actually see data from that implementation.

So, for example, with opportunity for improvement, if it is a new measure, they might submit something from the literature about how that particular focus of measurement is or is not being implemented or where the performance gap is. If it is a measure that has been implemented, we would like to see what the data are for that measure, how are the facilities or physicians doing on that measure.

One of the things, as you know, with opportunity for improvement, it is under

our importance to measure and report. Under the new guidance, all three of the subcriteria must be met.

One of the things -- and I will go into it in a little more detail in just a moment -- is about a potential for reserve status. If a measure has a high level of performance -- and this would only be for an already-endorsed measure -- if when it comes back it has a high level of performance, typically, that would not meet our criteria of opportunity for improvement. But we have implemented a process where in exceptional circumstances we can endorse a measure in reserve status, meaning it is a highly-credible, reliable, evidence-based measure currently with high levels of performance. And we would endorse it kind of if it is needed to be used. So, I will talk a little bit more about that in a minute.

The other thing is that reliability and validity for an endorsed measure, the goal

is, hopefully, that there has been some expansion of that testing, unless it is already at that high rating. But that is something that we will take a look at as we get into the individual measures.

Usability for an endorsed measure, again, we would actually like to get some information about use of the measure. Typically, an endorsed measure, when it comes back for endorsement maintenance, has been endorsed for up to three years, maybe a little less, maybe a little more. And so, we would like some information on how it is actually being used.

And then, feasibility, certainly if there has been any unintended consequences as a result of implementing a measure.

Okay, next slide.

So, you know that our rating scale is high, moderate, low, and then insufficient or insufficient evidence. For some of the criteria, we have a generic rating scale.

Then, for evidence reliability and validity, there are those very specific definitions for what constitutes high, moderate, and low.

Okay. The other thing, next, yes. We have had some questions, and we just want to clarify the difference between a low rating and insufficient evidence or insufficient information. Basically, a low rating generally means that the evidence or information demonstrates that the criterion is not met; whereas, insufficient evidence or insufficient information means the evidence does not exist or nothing was submitted or inadequate information was submitted.

And we like to keep that distinction that low means that, really, whatever was submitted demonstrates it was not met versus insufficient to make that determination. If the reason for the insufficient rating is because the measure developer didn't submit something, that can be, of course, remedied versus if it just

doesn't exist.

Okay, the next one.

So, then, that brings us to the question of how to deal with measure submissions that may have been inadequate versus the evidence just doesn't exist. And I know that this has come up on some of the discussions and questions and our optional call.

So, what we had suggested on your preliminary evaluations is that, if the information was not sufficient, the evidence wasn't provided, to go ahead and rate that as insufficient, but to make a note that you know that there is evidence that supports it or information or data that supports that particular criterion.

When we discuss the measures, after the Committee's discussion, if the Committee is really confident that that evidence does exist, you can rate the evidence based on your agreed understanding of the evidence, and we

could ask the measure developer to update their submission with evidence that does exist. And we could make the recommendation provisional on having that additional evidence.

It is kind of a fine line, and we will certainly rely on your expertise and judgment and assistance with this. But what we are hearing is that sometimes the measure submission may not totally represent the evidence that exists. And so, we do really rely on your knowledge and expertise.

But, as you all know, assimilating a body of evidence and grading it is a big project in and of itself. So, it is not something we expect you to do on the fly, but certainly some of you, many of you have been very intimately involved in evidence reviews and have knowledge of what has been assembled.

Okay, next one.

So, importance to measure and report, as we have mentioned, this is a must-

pass criterion. The recent guidance is that all three subcriteria must be met. This includes high-impact aspect of healthcare. This generally involves a large number of patients, high resource use, high consequences or severity of consequences of poor quality.

The next one is gap in performance or opportunity for improvement. In this one, we are asking for data about variability in performance or overall poor performance. That could be a situation where it is not really much variability, but it is just being done poorly across the board.

And then, the last one is evidence supports the measure focus. Particular health outcomes have an exception to having to present a body of evidence, but, certainly, a process measure, a structure measure, other types of measures should have an evidence base that says that is an effective intervention, service, treatment to warrant having a performance measure.

So, I think this was an area that we had a lot of discussion at our last meeting, and to really make the distinction that what we are talking about here is importance to measure and report, meaning it is important to have a performance measure that, hopefully, will be publicly reported at some point, versus everything that is important to do in day-to-day practice of renal patients.

So, I know that this will come up in some of the measures that we are going to be reviewing today, but I just wanted to, again, make that point that there are thousands of things that are important to do in practice of care. They don't all need to have a performance measure. And I am sure we will have some discussion about that as we go through the measures.

CO-CHAIR CROOKS: Karen?

DR. PACE: Yes?

CO-CHAIR CROOKS: I would just like

to interject a quick question. I think it is appropriate.

As we are going through this process, when we get into the evidence supports measure focus, there is a question, is this a health outcome or not, or I guess a process? And then, it says, if it is a process, then you apply that chart with the body of evidence.

DR. PACE: Right.

CO-CHAIR CROOKS: But if it is a health outcome, you don't? Is that right?

DR. PACE: Right. So, that is a good point, and I think we have had this -- it has been a little confusing. But, basically, if it is a health outcome, and there is a rationale for its relationship to healthcare services, then the developer does not have to present a body of evidence to support it because it is a health outcome. And there could be multiple bodies of evidence that relate to that health outcome.

However, if they do submit a body of evidence, I mean that strengthens the application and it strengthens the interest in that particular outcome. And we would ask you to at least rate it since they went to the trouble of providing that additional information.

So, when we go through individual measures, we will kind of skip that question about health outcome, is there a rationale? If it is not a health outcome, we don't need to talk about that particular --

CO-CHAIR CROOKS: Well, we can save some time, though. If it is really a health outcome, why should we review the evidence in this setting if we are going to be pressed for time?

One other comment, too. The other thing that seems clear, but sometimes can be fuzzy, what is a process and what is an outcome? Because some seem that it could be both or it is intermediate somewhere, and how

do you make a clear distinction there?

DR. PACE: Right. That is sometimes quite difficult. Health outcomes tend to be end result, you know, obviously, mortality, or some things that actually are proxies for health outcomes, such as readmission. Intermediate clinical outcomes tend to be more the things related to clinical parameters, such as the hemoglobin value, a lab value of some sort, the blood pressure level. And then, process generally is some type of treatment, service, intervention. But sometimes it is based on perspective.

CO-CHAIR CROOKS: Yes.

DR. PACE: And so, for some things it is much more clear and others not. But the exception for presenting a body of evidence is really just for the health outcome, not for intermediate clinical outcomes. Okay?

CO-CHAIR CROOKS: Thank you.

DR. PACE: All right.

Okay, so next slide.

Opportunity for improvement. So, as I said, we are looking for variability or overall poor performance. This is where I am going to talk about potential reserve status for endorsed measures that have demonstrated high levels of performance. And actually, I think the first measure we get into is one that fits that category.

And the reason for implementing this reserve status is to retain endorsement of reliable and valid quality performance measures that have overall high levels of performance, so that the measure could be used in the future, if necessary.

This is intended to be for an exceptional circumstance, not the rule for every endorsed measure. And there are certain criteria that would apply.

One of the things that we really want to focus on is high levels of performance that are actually due to quality improvement and actions to improve care versus problems

with the measure. So that I guess the best example of the distinction I am thinking of is in the past we had some measures about smoking cessation counseling, and the measure could be fulfilled, and people were concerned about it turning into kind of a checkbox type of measure. So that there were really high levels of performance, but most people thought it was really related to documentation versus any real change in care.

So, we will have to work through that, but the idea is that the high levels of performance improvement, hopefully, are really demonstrating that people have actually improved and are doing well with that particular aspect of care.

Okay. So, the next slide, then, is the criteria for reserve status.

When we are looking at performance gap or opportunity for improvement and the data that are submitted, we need you to kind of discern the distribution of the performance

scores, how many entities and patients does that include. So, if it is a very small sample, then maybe it is not really representative of what is really going on in the field.

Data on disparities, if they have provided data on disparities, so the overall scores could look fairly good. But if you start looking at disparities, there may still be opportunities for improvement in terms of disparities in care.

Again, size of the population at risk, the effectiveness of the intervention. So, all of these things need to be factored in, first of all, to your consideration of whether there is really high performance.

And then, the additional criteria that need to be met. So, the point we are bringing up is that, typically, if a measure doesn't have an opportunity for improvement, it won't pass importance to measure and report. We will stop at that point if it is

an endorsed measure and see if you think it is something that would need to be considered for reserve status, and we will continue the evaluation.

Lisa, do you have a question?

DR. LATTS: Yes. When you say "data on disparities," do you mean that there is data that there are no disparities within that measure?

DR. PACE: Right.

DR. LATTS: So, even if performance is high overall, but there is evidence that there is a disparity, we would not keep it --

DR. PACE: Right.

DR. LATTS: -- we would not put it on reserve --

DR. PACE: Right. Exactly.

Okay. So, next.

So, in addition to those considerations, criteria for reserve status is it would be measures that have strong direct evidence of a link to desired health outcomes.

Basically, we would want this for measures that are measuring things that are proximal to the desired outcome.

So, those two first bullet points lead us to generally measures that are more distal to the desired outcome would not be eligible for reserve status.

The reliability and validity ratings, our guidance is that they should be at the high rating. It may be too soon to be that stringent there, but we will have a discussion about that.

And again, as I mentioned, the reason for the high levels of performance is actually better performance, and that we hope to see demonstrated usefulness for improving quality and demonstrated use of the measure. Okay.

DR. WELCH: Karen, can I ask a question?

DR. PACE: Yes, uh-hum.

DR. WELCH: Can you go back and

talk a little bit about disparities again?

DR. PACE: Yes.

DR. WELCH: So, when you are talking about disparities, I think what I heard was the distinction between disparities in care versus disparities found within the measure. Is that what I heard?

DR. PACE: Right. So, for example, if we are talking about a measure of some care process, and if you look at the scores across the facilities, it looks like there is a fairly high level of performance across all facilities. But if you looked at that data in terms of differences between races, whether they get that particular process or service, that you may see some gaps there that didn't show up in the kind of higher-level analysis. So, that is what we are getting at there.

Does that make sense? Okay. All right.

Okay. So, next, we wanted to talk a little bit about the evidence guidance. And

some of the key points here from the Task Force were they really wanted developers to submit information so that the evidence that does or does not exist to support a performance measure is transparent to the Steering Committee, the public who review these measures, ultimately, the NQF membership that votes on them.

And the requirements are the requests for information about evidence is probably the biggest change in our measure submission form because the Task Force really wanted to get information about the quantity, the quality, and the consistency of the body of evidence.

The other key point is that the Task Force guidance was not that measure developers should be conducting primary evidence reviews. Obviously, that is a whole big endeavor in itself, but, hopefully, should use existing evidence reviews that have been systematically assembled and graded.

The Task Force identified that preferred evidence grading systems were U.S. Preventive Services Task Force and grade, but recognized that in today's current environment there are others or modifications. And so, those are acceptable. We ask which system them are using. And if it is other, that is fine to say "other", but to describe that. Expert opinion is not considered evidence.

We already talked about the exception for health outcomes.

Obviously, there is no kind of cut-and-dried way to do this. We still rely on the expertise and judgment of the Steering Committee.

Yes?

DR. FISCHER: With some measures that may be more novel, and there aren't reviews or body of evidence, then was it expected that there would be more of a literature review of individual studies or is that not what was expected?

DR. PACE: Well, it is a good question. I guess the answer would be, yes, if there wasn't any kind of body of evidence review that was existing to provide what evidence there was. And then, the Steering Committee would have to weigh that. We will get into that to a certain extent.

As we go through this, the experience that we are having with the Steering Committees as we are kind of trying to implement this Task Force guidance, we will try to have some debrief time at the end of our day tomorrow, so that we can get some of your feedback of what is working and what is not, and where we need to think about going back to the Task Force and making some revisions to that.

Okay. So, quantity, quality, and consistency of the body of evidence, and you know that there are specific rating scales describing these in the high, moderate, low, or insufficient evidence. And quantity is

simply how many studies in the body of evidence. Quality is about the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence. And there is a description of what kind of things are considered here. Certainly, the study design, flaws in the study, those kinds of things, directness of evidence. Then, consistency has to do with both the direction and magnitude of effects across the body of studies.

Okay, next slide.

And so, we have two exceptions to this. We have already talked about health outcomes. We really just need to see that there is a rationale for connection between the health outcome and at least one structure, process, service, intervention for healthcare.

Then, there is another exception, potential exception if there really is not a body of evidence, and about expert opinion. We will try to address that once we have

reviewed the body of evidence. And if something is not going to pass because of evidence, if the Steering Committee thinks that it is something that should continue to go forward, based on expert opinion, we will have a discussion about that.

I think one of the things that we are going to ask you to do, as we did last time, is that we really need to try to ground your recommendations in the criteria. If something doesn't meet criteria, but there is a reason for potentially continuing on with the recommendation, we have to have that well-documented for our reviewers and, ultimately, for our Consensus Standards Approval Committee and the Board, because they are really looking to see why measures that may appear not to meet our criteria are being recommended.

All right. So, let's talk about scientific acceptability of measure properties. Basically, this comes down to reliability and validity. Reliability

includes precise specifications. And generally, if that is a problem, that can be remedied by going back to the developer and asking that they define something or provide more specification. But we do expect reliability testing to demonstrate reliability at either the level of the data elements that are used in constructing the measure or the precision of the performance measure score. And I will talk a little bit more about that.

Validity starts with, is the measure consistent with the evidence that was provided in support of the measure? And then, validity testing. Finally, if there are potential threats to validity, has an analysis been done to really demonstrate that those are resolved or not a problem? So, this is an issue, if it is an outcome measure risk adjustment, is the risk adjustment adequate? If there are exclusions, are those exclusions justified either by the evidence or, if they are not clinical evidence-based exclusions,

what is the analysis of what effect those exclusions have on the measure?

An lastly, we have disparities. You know, the measure being specified for disparities under 2c, we will look at that after reliability and validity. We are still working with where that best fits, but the Measure Testing Task Force decision logic really applies to reliability and validity.

Okay. So, next slide.

The Measure Testing Guidance, the key points from the Task Fore were, first of all, that reliability and validity should be demonstrated through empirical evidence. It is not something that you say we agree it's reliable or we agree it's valid. It really should be demonstrated through empirical evidence.

And the other thing to keep in mind is that reliability and validity are about the measure as specified, not about some concept in the literature. It really is a measure

property. So, it has to be the measure that is being presented is what was tested for reliability and validity.

The Task Force really tried to provide flexible testing options rather than being prescriptive because there's a lot of factors that go into choosing a particular method. They did not set specific thresholds, again, because they thought that there are too many factors to have a hard-and-fast threshold of what the reliability statistics should be, for example.

Here in testing, insufficient evidence basically means it wasn't tested. Perhaps it could be that they just didn't provide the information we needed, and we can get that clarified. Again, we still need expertise in judgment to look at these.

The Task Force also came up with some strategies to mitigate the burden of testing because they know that testing does require resources. So, the rating scale is

really based on allowing measure developers to test reliability and validity for either the data elements that go into constructing the measure or the precision or the measure's score itself.

So, when we talk about the data elements, we are really talking about the data that are captured and used in the measure. If it is a diagnosis, that is one part of a measure. That is one data element for a measure. Age might be a data element. Whether an intervention was provided is a data element versus the actual computed measure score, which might be the percent of patients who had "X" or percent of patients who died, et cetera.

The Task Force did say that testing could be done on a sample and that, if empirical evidence of the validity of the data element was done, reliability of the data element at that level would not need to be addressed. And I know that this is very

technical and complicated, but data element validity is really focused on is it correct data. And so, it is really comparing it to some authoritative source. Whereas, reliability of the data element is about repeatability, reproducibility.

So, if you are doing medical record abstraction and you compare results of two abstractors, that would be a reliability test. If you were looking at claims data and comparing it to what is in the medical record, that would be more of a validity testing. Is the information on the claims actually an accurate representation of the medical record? So, we will get into these nuances with some of the measures as we go through them.

Yes, go ahead.

CO-CHAIR CROOKS: I think this is a good time to ask this question.

DR. PACE: Okay. Good.

CO-CHAIR CROOKS: Several times we saw, or I saw in my work, that if it is

electronically-submitted data, that it is probably reliable. As you just were describing what reliability is, if it goes in and it is stored, it should come out the same way.

DR. PACE: Right.

CO-CHAIR CROOKS: But there is a still a burden to do validity testing, and maybe some validity testing at the data element level. Am I correct?

DR. PACE: Right. Exactly. So, the Testing Task Force Report did some work specifically on electronic health record data, which I think some of that also transfers to claims data. So, if you have data in an electronic database and you apply your computer program, you are going to get the same answer. It is going to be repeatable. It is going to be reliable.

But keep in mind that just because it is repeatable doesn't mean it was accurate or the right information in the first place.

And so, that is the difference between validity, and, basically, if you are relying on data-element-level testing, there is a way to do some validity testing of electronic data as well.

Lisa?

DR. LATTS: Yes, I agree that it is an important standard. I just worry that if we are too stringent with it, it would mean that essentially all of our claims-based measures would have to be thrown out the window because, you know, depending on your perspective, I think claims are probably not that accurate. Yet, it is sort of a standard that we have today.

DR. PACE: Well, I think that is the question that we are asking everyone to address. The criteria actually allow face validity of the measure score. And so, that is a pretty weak requirement. And we allow face validity if it has been systematically assessed, not just somebody says, "We agreed

this was a valid representation." But it is a question that comes up: are the data valid?

And again, the Task Force Guidance is to allow testing at either the data element or the measure score. So, if people feel it is too burdensome to look at the data elements, then they are going to have to rely on hopefully doing some empirical validity testing. But, currently, our criteria allow for face validity.

So, there are multiple ways that people can address this. They have to make the case, first of all, to you, as the Steering Committee, and then, ultimately, beyond this. We will see some of that, as you know, with some of the measures that come up, and we will kind of work our way through that.

But the criteria apply equally to measures regardless of the data source. Reliability and validity are kind of basic principles of measurement that our Task Force and CSAC and Board feel apply to all measures.

Okay. The other thing is that prior evidence could be submitted. So, for example, on the claims question this comes up probably more often. But if there was some study done that showed that claims for a particular diagnosis are highly reliable, the Task Force said go ahead and submit that as evidence of reliability or validity for that particular data element.

The other thing is that, just in terms of the question about claims data, it may be known for certain diagnoses that claims data are highly valid and accurate versus other types of diagnoses or procedures. So, again, you cannot make an across-the-board assumption. It depends on what you are trying to extract, what concepts you are measuring.

Okay. And we will continue on. So, reliability testing at the data element level, that the data elements are repeatable, producing the same results a high proportion of the time when assessed in the same

population in the same time period and/or that the measure score is precise.

So, at the data element level, the key question is, has it been demonstrated that the data captured or used in the measure are repeatable and reproducible? And we have already talked about at the data element level, if you have done validity, then you don't have to do separate reliability.

And at the measure score, has it been demonstrated that variability across entities is due to true difference or signal versus error or noise? So, again, I think when we get into some specific measures, we can look at these examples a little bit better. But certainly, if there are any questions, we can talk about those now.

Yes?

MS. ANDERSON: I have a question. Many of the data elements, at least in those that I reviewed, were coming out of CROWNWeb. If CROWNWeb doesn't have those data elements

currently, it was very difficult to figure out how do we evaluate those measures.

The other concern I have is, for those who are doing manual data entry, the inaccuracies of the data is pretty high. So, how do we address that?

DR. PACE: Right. Well, the CMS measures, they did test with actual CROWNWeb data. I know that there are a few other measures that were submitted saying that eventually those were going to be in CROWNWeb. And I think that is a discussion that you all will have to have. If they are not currently in CROWNWeb, so we don't have those specifications, what would you be recommending for endorsement?

So, I think we just have to have a discussion about that. I know that is one of the questions we asked for clarification for some of the measure developers. Certainly, they are all here to respond to questions as they come up from the Committee.

DR. WELCH: Can I ask one more question?

DR. PACE: Yes.

DR. WELCH: It has to do with reliability. Did I hear you say that data that is collected using an electronic health record is reliable? Because I am thinking of all the error that is involved with entry of that data and collection of that data before you a score.

DR. PACE: Right. And this gets into, I think that points to the distinction between reliability and validity of the actual data. So, reliability is about reproducibility. So, if you have a software program that has been designed to pull the data used for the measure out of electronic health records, if you run that once and then run that again, you are going to get the same result. So, it is reliable or repeatable.

The key question is whether it is accurate. And that is a validity question.

So, is the data that is being pulled for this measure, is it an accurate reflection of the real data? So, for example, maybe it is pulling from the wrong field, but that is a validity question. And there are ways to look at validity of data from electronic health records. Again, it involves some comparison and abstraction.

But I think if you keep in mind at the data element level reliability is about repeatability, reproducibility, versus validity is more about is it the correct data; is it the accurate reflection of the data. Okay?

All right. Okay. So, validity testing, again, could be done at the data element level, as we were just talking about, or at the measure score level.

Has it been demonstrated that correct and accurate conclusions about quality can be made when we are talking about the measure score? And again, we do allow face

validity. Again, we ask that they have something that somehow they have systematically assessed that, whether that is actually having a group of experts vote on that. Again, it is about the measure as specified. Will those scores actually reflect the level of quality in a particular facility?

Actually, in some of the measures we have seen really nice validity testing, you know, looking at the conceptual relationships between a process measure and an associated outcome measure. That is ultimately what we would like to see, that if we measure performance on a particular process, how is that reflected in what we are actually trying to achieve with patients?

Okay. So, in your evaluation of testing, we ask you to consider was an appropriate method used. I know this gets a little tricky, but we need to consider the level of testing of the data or score, what data source was used, the type of measure, the

topic.

I mean there is a variety of things that go into selecting a method for reliability testing. We have asked on the submission form for developers to provide a rationale for that. In some cases, that wasn't provided and it wasn't clear what was being submitted or why that was considered a test of reliability or validity.

We posed those questions to the developers. I think everyone received all the responses we have. I think some of that was cleared up.

Was the scope of testing adequate? So, if it is a sample, consider the number of measured entities, facilities, physicians, the number of patients, and the representativeness of who was included in that sample

And then, ultimately, were the results that they obtained from their testing actually within norms and demonstrating that we have a reliable or valid measure?

Okay. I know I have gone over here, but I think it helps to get us kind of squared away here before we get into actual measures.

So, usability is the extent to which intended audiences can understand the results of the measure and are likely to find them useful for decisionmaking. We have subcriteria about public reporting and quality improvement.

We in this go-round have really asked for a rationale. Occasionally, measure developers have actually done some testing with their audiences on this. Of course, if they have, that is great, and we ask for that information. But, primarily, I think you will see a rationale in those sections.

As I mentioned earlier, for endorsed measures, we actually would like to see that they are actually being used. And so, we will be looking at that as well.

Feasibility, hopefully, the data

for clinical measures are actually data that are being used in providing care.

Electronic sources of data certainly are less burdensome. Hopefully, there is some discussion about potential susceptibility to inaccuracies or unintended consequences, and that the data collection strategy can be implemented.

Related and competing measures, we are not going to address in our first go-through of the measures. This is something that we will look at. At the end of tomorrow, if we have related and competing measures, then we will talk about how we are going to address those going forward.

Okay. So, I am going to just quickly go through what we are going to do today at the meeting.

We will have periods for NQF member and public comment twice each day. We are going to have the measure developers briefly introduce their measures at the beginning of

each topic area. We ask the developers to really keep this two to three minutes because they have presented the information to you in the measure submission about their measures. They are here and available to respond to questions from the Committee as needed.

The Steering Committee will discuss and vote on each measure, and we will do that by criteria. I will be introducing the first measure, but we will try to go through each criteria that we are going to ask you to vote on before we move onto the next one. That way, that discussion will be fresh. We can vote and then move on to the next one.

So, what our process will be is to have one Committee member begin the discussion. We have asked you to summarize the preliminary evaluations, really identifying where there were questions, concerns, or differences of opinion, so that we can really focus on the things that need to be discussed.

Then, we will ask the other assigned reviewers if there are any additional comments, and then have full Committee discussion, followed by a voting on that particular criterion.

As Helen mentioned earlier, and we handed out that four-pager kind of quick reference, if a measure does not pass importance to measure and report, we will stop there. The same way, if it doesn't pass scientific acceptability, we will stop there.

Okay. Your votes today will really be conditional on whether there are related and competing measures. So, we ask at the end, overall, has it met the criteria? But if there are related measures where there may be measure harmonization issues or competing issues, two measures, basically, on the same issue, the vote is not final until those issues are resolved. So, this is kind of your preliminary.

Okay. And, Lauren, do you want to

just mention -- we won't do the example, but do you want to just mention the electronic voting?

MS. RICHIE: Everyone should have a remote control. Does anyone not have one? Okay, I just want to make sure we all have one.

Just like before, the criteria will be reflected on this screen here and we will vote according to -- and we will go through a sample. Once we go through the first measure, we will just do a test to make sure all the remotes are working.

And you also have a one-page instructional sheet. It should have been underneath your agenda, just a quick instruction on how to use the remotes, but it is fairly simple. We will go through a sample.

You will have up to 60 seconds to vote. Then, you will just press a number on the keypad that corresponds to the response

that is on the screen there. You don't have to hit the Send key. You can just do 1, 2, 3, et cetera. If you want to change your answer, you just change it to the number that you want to use. So, 1 to 2, and you don't have to hit Send after that.

Then, after everyone has voted, we will see a tally on the screen of the number of votes as well as the percentages.

All right. And I believe you are ready for the first measure.

DR. PACE: So, first, we will be starting with the anemia measures. And so, you're right, we will start with the measure developer --

CO-CHAIR CROOKS: So, the measure developer for 252.

DR. PACE: I know, but who are they though? So, CMS and PCPI.

Okay. So, how about does CMS want to briefly present your measures?

MR. WOLFE: Thank you very much. I

would like to thank the Committee for their time and their expertise.

DR. PACE: Could you tell us who you are, just for the record?

MR. WOLFE: To the NQF organizers, I am Bob Wolfe. I am with Arbor Research, which is the contractor for CMS. Actually, CMS is the measure steward and developer.

I would like to just say a few comments about the anemia measures that we have. There's one process measure having to do with ferritin and three target measures. I want to say a bit about each one.

The process measure we have prepared, and there were questions about or issues perhaps related to performance gaps, which are correct. If you look at the data, you will see that most facilities are performing quite well on this measure.

We are coming into a time of bundling. It is very plausible that there will be incentive changes and practices. I

would ask the Committee to consider that in thinking about whether performance gaps from the past are the only relevant issue in evaluating the importance of a measure.

For the target measures having to do with hemoglobin levels, the level of evidence from clinical trials is relatively clear for hemoglobins greater than 12. There is very little evidence about hemoglobin levels less than 10 from the clinical trial data. But the hemoglobin less than 10 has been withdrawn by CMS. So, that one isn't relevant.

For some measures in general, I heard there were comments about the validity and reliability of the data. I would like to say that most, if not all, of the CMS measures that you will hear about during these two days are based upon facility-level measures. While it is very valuable if every single patient number can be reported correctly, in fact, at the facility level with an adequate number of

either patients and/or time of followup, individual errors at the patient level are overcome because we are reporting an average. So, it is really the reliability of the measure at the facility level that is important rather than only at the patient level.

And again, I would ask the Committee to consider that as you are thinking about the actual use of the measure. I know many doctors think about it at the patient level, and that is very important. But it is actually being used at the facility level, where some inconsistency in reporting at the patient level can be overcome by data.

Thank you very much.

CO-CHAIR CROOKS: PCPI also submitted anemia measures.

MR. JONES: Sure. Thank you.

Ed Jones. I'm a member of the KDI Work Group from the AMA PCPI and, along with Barbara Fivush, will be representing that

group as the developer of the measures.

I am going to really just get into the approach that we have for measure development and testing and would rather leave the questions for the specific measures when they come up.

The PCPI combined CKD Work Groups along with Adult and Pediatric ESRD Groups in order to identify and define quality measures toward managing and improving outcomes for our patients.

Two of the measures dealing with adequacy were previously endorsed by the NKF and, therefore, are up for review for maintenance. I want to point out specifically that all eight of the measures have been tested for reliability and validity, and seven of the eight are being used in CMS's PQRS Program.

The measures were developed through a rigorous, evidence-based process that has been refined and standardized for over a

decade. Measures are developed through a cross-specialty, multidisciplinary group, including nephrologists from the RPA, ASPN, and NKF, along with endocrinologists, methodologists, internists, preventive medicine, and family doctors.

Practice guidelines are used as the foundation for the development of the performance measures, and the guidelines with the strongest recommendations and with the highest level of evidence are being used.

The Work Group reviewed available information in gaps of care and unexplained variations in care to ensure that the measures represent areas most in need of performance improvement.

The Work Group also reviewed data regarding feasibility, reliability, and exception reporting available from implementation of a subset of 2007 measures. The Work Group made every effort to harmonize these performance measures with similar

metrics.

It is important to recognize that outside stakeholders are involved, in particular, those clinicians who will implement the measures. Therefore, all measures were released for 30-day public comment and were peer-reviewed. All comments were reviewed by the Work Group and modifications were accepted as appropriate.

The measures submitted for your consideration are specified to ensure widespread implementation using EHR when possible.

In summary, the Work Group sought to focus on those areas with the most potential for impact, where there was the strongest consensus about the best practice, and where the likelihood of unintended harm was the lowest. Moreover, the group sought as much as possible to keep the measures straightforward; aligned, when appropriate, with measures developed by others, and

clinically-sensible, giving the clinician the latitude for judgment about the appropriateness of the intervention.

And again, we will address specifics of the measures during the question period.

Thank you.

CO-CHAIR CROOKS: Thank you.

Okay. Then, let's go with 252.

DR. PACE: Okay. All right.

So, I just want to do one clarification, though, with Bob. CMS withdrew their hemoglobin target measures.

MR. WOLFE: Thank you. And I said that very obliquely when I said that CMS had withdrawn them.

DR. PACE: Okay.

MR. WOLFE: So, I want to clarify that those are not CMS-sponsored. Thank you.

DR. PACE: Right. Okay. Thank you.

Okay. So, we are going to start

with Measure 0252, which is assessment of iron stores.

Before we get into this measure, we just want to mention that we can display information from the measure submission form, if that is needed, or any of the documents. Also, we will be looking at the tally of the preliminary results.

I know that most of you brought your own files. But if you need something or you want something displayed, certainly let us know.

So, what I will do is do a summary of the preliminary evaluations, and then we will vote on each of the subcriteria or criteria that we need to as we go through them, so that it is fresh in our minds.

Okay. So, with this measure, I will start with just a brief description. This is a measure of the percentage of all adults greater than or equal to an 18-year-old hemodialysis or peritoneal dialysis patient

prescribed an ESA at any time during the study period or who have had a hemoglobin less than 11 in at least one month of the study period, for whom serum ferritin concentration and either percent transfer and saturation of a reticular site hemoglobin content are measured at least once in a three-month period, for in-center hemodialysis patients, peritoneal dialysis patients, and home hemodialysis patients.

So, the other thing to note is that this is a measure that is up for endorsement maintenance. It was endorsed in November of 2007.

The other thing I want to note that in our just-now-completing ESRD Project, those of you who are on the Committee, I just want to refresh your memory that we had two measures submitted on assessment of iron stores in the ESRD Project which this Committee did not recommend go forward.

So, one of those measures submitted

in the last project was intended to be a replacement for this one. We can talk about where I think they had removed measuring the reticular site hemoglobin content.

But this is the measure that was endorsed. It is up for endorsement maintenance.

I just wanted to remind people of where things were at with those last measures. And the primary reason that the measures were not recommended in the last project was the issue of assessment measures being distal to the desired outcome. At that time, we actually had a measure of hemoglobin value.

If you remember in that project for the pediatric and, also, in this project, initially, we had an endorsed measure for hemoglobin values less than 10. We will get into this in a moment. I am just trying to put the context with this measure.

For the pediatric measure in the ESRD Project, CMS at the end, even though that

measure went through our whole process, withdrew that measure due to the recent FDA announcement that came out. We have provided all of that information to you all.

So, the pediatric hemoglobin-less-than-10 measure will not be endorsed, and CMS is also withdrawing their previously-endorsed adult measure of hemoglobin less than 10.

So, I mention that because that was part of the issue of not endorsing an assessment measure about iron stores because, ultimately, we have the hemoglobin value, and if there are problems with hemoglobin, one of the responses, obviously, is to look at iron stores.

But I will go on from there. I just wanted to kind of remind people where we were with those measures.

Okay. So, the first subcriterion that we ask you to vote on is on high impact. Basically, the preliminary reviewers all agreed that this was either a high or moderate

rating. So, I will just see if there is any additional comments or issues that anyone wants to bring up. Otherwise, we could actually vote on this subcriterion and move on.

So, any issues about high-impact aspect of healthcare for iron stores in ESRD patients?

(No response.)

Okay. Now we didn't do a practice round with you, but for those of you who were here last time, you know. But what we just want to remind you is that your vote will not register until it actually is a timer that starts. You will have up to 60 seconds to enter your rating. So, you will be using a number between one and four for this.

CO-CHAIR CROOKS: Do you need to press Send to get your vote to go?

DR. PACE: No, you do not need to press Send. If you want to change your vote, you can change your vote before we stop the

timer. You would change your vote by just simply pressing another number. That is all you would need to do. It will capture your last number. Once we know that everyone's vote is in, we will go ahead and stop the timer.

Lauren will be asking Kristine and Lorien on the phone to give us their vote, so Lauren can enter that.

CO-CHAIR CROOKS: Is Kristine with us yet?

CO-CHAIR SCHONDER: Yes, I am.

CO-CHAIR CROOKS: Oh, hi. Okay.

CO-CHAIR SCHONDER: Hi.

CO-CHAIR CROOKS: Do we need to have her do her introduction?

DR. PACE: Yes, I guess we should, yes.

CO-CHAIR CROOKS: Yes. And Andrew isn't here yet, is he, Dr. Narva?

DR. PACE: No.

CO-CHAIR CROOKS: Okay. Kristine,

will you please do your introductions briefly, name, position, city you live in, a bit about your quality background? We have been giving one brief non-medical interest and then your disclosures.

CO-CHAIR SCHONDER: Okay. I am Kristine Schonder. I am a clinical pharmacist with the Starzl Transplant Institute in Pittsburgh, Pennsylvania.

My background with quality improvement is I have actually worked on all three of the Renal Committees here for the NQF and co-chaired the last Steering Committee with Peter as well.

My non-medical interest, actually, I'm on the beach and away from all of you. That is why I am not there, but I am on vacation with my children. It is our first vacation since we have adopted two children from Russia. So, sorry I couldn't be there with you, but you know the beach is actually calling a little bit more right now.

And as far as my disclosures, I do have no disclosures.

CO-CHAIR CROOKS: Okay. Thank you, Kristine.

So, we can go ahead with the --

DR. PACE: Right. So, unless anyone has anything to add or bring up about high impact, we can --

DR. DALRYMPLE: Karen, can I just ask a quick question? This is Lorien.

DR. PACE: Yes.

DR. DALRYMPLE: The one through four, what does that correspond to? I'm sorry, I don't see that on the slides. I don't have a keypad.

DR. PACE: We are voting on impact under importance to measure and report, and the options are high, moderate, low, or insufficient.

CO-CHAIR CROOKS: Right. So, when called, you can just respond on that scale.

DR. DALRYMPLE: I'll just give

answers instead of the one through four. Is that okay?

DR. PACE: I'm sorry, what?

DR. DALRYMPLE: So, I can just respond as high, moderate, low, insufficient as opposed --

DR. PACE: Yes, yes, please use the words instead of the numbers, so we make sure we are getting it correct.

Right, the rating scale is also on the submission form and Excel file, et cetera, but, yes, definitely on the phone, just use the words high, moderate, low, or insufficient.

CO-CHAIR CROOKS: I think in terms of process, let's us vote first and then we will poll them.

DR. PACE: Okay. All right.

CO-CHAIR CROOKS: Okay.

DR. PACE: We want to do it, actually --

CO-CHAIR CROOKS: Otherwise, they

will have undue influence on us.

DR. PACE: Well, Lauren wants to be able to do it, enter their votes electronically.

CO-CHAIR CROOKS: Oh.

DR. PACE: So, we will go ahead and start. I mean it would be easier to ask them first, right?

MS. RICHIE: Yes.

CO-CHAIR CROOKS: You would rather ask them first?

DR. PACE: Yes.

CO-CHAIR CROOKS: Okay. Everybody, don't listen.

(Laughter.)

Okay. So, you go ahead.

MS. RICHIE: This is Lauren.

Lorien, I am going to take your vote now for high impact. So, high, moderate, low, or insufficient?

DR. DALRYMPLE: For impact, moderate.

DR. PACE: Yes, start the timer. And you can all vote now. Once the timer starts --

CO-CHAIR CROOKS: Oh, we can do it simultaneously?

DR. PACE: Yes, you can do it simultaneously. Okay.

MS. RICHIE: And, then, Kristine, high, moderate --

CO-CHAIR SCHONDER: Moderate.

DR. PACE: Okay, we're going to start over. We need to have 60 seconds.

DR. BERNS: Can I just ask a question and clarify here?

DR. PACE: Yes.

DR. BERNS: When we are evaluating this, it is really on the importance of reporting as much or more than the importance of measuring. Measuring is just routine care potentially, as opposed to the importance of reporting this publicly.

DR. PACE: It's really both. It's

really both.

DR. BURSTIN: It's really the importance of the performance measure itself.

DR. PACE: Right.

DR. BURSTIN: So, not just the routine clinical care, but the actual use of the measure for quality improvement as well as accountability functions.

DR. PACE: All right, so we are starting over on -- no. We'll see if this gets working. Otherwise, we will take our quick break, and then we will get this settled.

With the people on the phone, we are having trouble getting them. We could add them at the end.

Okay, let's take just a 10-minute break a little bit early. We'll get this fixed, and then we will be able to proceed more smoothly.

(Whereupon, the above-entitled matter went off the record at 10:44 a.m. and

resumed at 10:59 a.m.)

CO-CHAIR CROOKS: Okay, to clarify the voting procedure, Lauren has a clicker for Lorien and a clicker for Kristine. So, that is what happening here. So, we will hopefully have one minute, and during that time she will poll them and we will all vote, and it should work out just fine. So, that is the process we are going to try.

DR. PACE: Okay. So, we are back to impact, and this is Measure 0252, assessment of iron stores. We are on impact.

Tenee, go ahead and start the timer.

MS. RICHIE: And, Lorien, you said two, moderate, correct?

DR. DALRYMPLE: Moderate, correct.

MS. RICHIE: Okay. And, Kristine, moderate, correct?

CO-CHAIR SCHONDER: Yes.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: So, we are still

missing some 12 --

DR. PACE: One. We have 21 people.

CO-CHAIR CROOKS: Oh, 21. Twenty-one, okay. Just one person. You might vote again because it doesn't hurt.

DR. PACE: Yes.

CO-CHAIR CROOKS: To make sure your vote went through.

DR. PACE: Okay. There you go. All right, there we go. Okay. So, that's how it works, everybody.

CO-CHAIR CROOKS: And we need to read it into the record, right?

DR. PACE: Yes.

CO-CHAIR CROOKS: High, 5; moderate, 14; low, 2; nobody voted insufficient.

DR. PACE: Okay. So, I will move on to 1b, which is opportunity for improvement. Basically, the preliminary evaluations were split between either moderate or low.

And so, one of the things I will just mention, in the measure submission -- and, Lauren, you may want to bring this up, 1b2 -- they provided an analysis of CROWNWeb data. The distribution of scores at the first quartile was 97 percent; the median was 100 percent, and the third quartile, 100 percent. So, there really is very little variability, and, overall, it looks like pretty high performance.

They did present some information in 1b4 about data on disparities by different population groups. Those also look to be quite high.

So, that is, I think, the information on opportunity for improvement. But I will stop there and ask for the other assigned reviewers, if they want to make any comments on this, and then whether there is any discussion.

One of the reviewers noted that, yes, it is high performance; perhaps that

indicates that the measure was working in terms of getting people to high performance.

So, I will just ask if there are any other -- the assigned reviewers, and then a general discussion. Anything else from the assigned reviewers, additional comments about that?

(No response.)

CO-CHAIR CROOKS: Is it okay to tell the panel or would they like to know how the other reviewers voted on this?

DR. PACE: Yes, and those are --

CO-CHAIR CROOKS: Oh, you can see it? Okay.

DR. PACE: Right, right, right.

CO-CHAIR CROOKS: Amazing. Okay.

(Laughter.)

DR. PACE: So, the preliminary voters: two, moderate; three, low.

CO-CHAIR CROOKS: Okay.

DR. PACE: Okay. Any discussion?

(No response.)

Okay. Then, opportunity for improvement or performance gap, again, the scale is high, moderate, low, insufficient. And go ahead and start the timer.

MS. RICHIE: And, Lorien, performance gap, high, moderate, low, or insufficient?

DR. DALRYMPLE: Low.

MS. RICHIE: And, Kristine?

CO-CHAIR SCHONDER: Low.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: So, the results are 16, low; 5, moderate.

DR. PACE: Okay. So, even though all of the subcriteria need to be passed, we are going to rate all the subcriteria, and then we will apply the decision logic to importance and then discuss any potential for reserve status.

So, the next subcriterion is about evidence. This is not an outcome measure, so we will skip that particular question.

And we will be talking about the quantity, quality, and consistency of the evidence. We have to look at those separately.

On the quantity of evidence, the preliminary voters agreed that this was moderate or high.

And maybe what we'll do, it might be easier to kind of talk about evidence in general and then rate the specific subcriterion on evidence. See what you think about this, but it might be a little too disjointed to just vote on those separately.

So, quantity is obviously just about the number of studies. Our scale is pretty generous. The low would be one study; moderate, two to four, and high, five.

So, I think the main issue that was brought up -- sorry. In the quality of evidence, we had two preliminary reviewers that indicated insufficient; two, moderate, and one, high.

We did ask the developer for some additional clarifications about the evidence, and that was in some of their responses.

I think the thing to consider here is, you know, again, we are measuring something that is distal to the desired outcome. It is about assessing iron stores. The evidence in the additional information that was submitted talked about the evidence of treatment using iron and the ability to have lower ESA doses if you treat anemia with iron. Also, of course, they had mentioned the association between hemoglobin levels and mortality.

So, again, this is one of those things where you assess and then you diagnose, identify treatment options, administer treatment, and then effect on hemoglobin, and then, ultimately, on mortality or survival.

So, the evidence would be indirect because there is not evidence about how frequently you should measure hemoglobin. It

is one of those things that is necessary, but not sufficient to achieve the desired outcomes.

As I mentioned earlier, this was originally in the context, and, also, when we looked at this in the prior project in the context of having a hemoglobin target measure or at least measuring patients below 10.

So, I am going to stop there and ask the other reviewers to make some comments about evidence in general, and then open it up to the Steering Committee. Then, we will actually vote on quantity, quality, and consistency.

So, some of the other reviewers -- Lorien --

DR. DALRYMPLE: Yes. Hi.

So, I was one of the reviewers who in initially thought there was insufficient data on quality and consistency. However, supplemental data, and I think made available to everyone, provided I think further

substantiation of the quality and consistency of data. And they provided additional references.

I am not sure on some of that additional data if everybody has had a chance to look at it, but I think it is relevant to the criteria.

DR. PACE: Right. And, Andrew Fenves, Rick Kaskel, Kristine, any additional comments?

CO-CHAIR SCHONDER: No, I agree with Lorien. That was my feeling as well. I originally rated it as insufficient evidence, but I think the additional evidence that we received substantiates the measure better.

DR. PACE: Okay. So, could we have some discussion, then, from the Steering Committee in general about the evidence that supports this measure.

CO-CHAIR CROOKS: Well, this is a question I have, and this applies, I think, to many of the things we are going to look at.

The evidence is that it is good to treat anemia with iron, and that's a good thing. I don't think there is much doubt about that among those of us in practice, anyway.

But there is no evidence saying that to measure the frequency or to make sure that the measurement is done with a certain frequency gets us there that I read. And I have to confess, I didn't see the stuff that was submitted later. But does that bother people? Is that what we are supposed to --

DR. PACE: Right.

CO-CHAIR CROOKS: We have to make that leap or -- it bothers me.

DR. BERNS: Having just reviewed this for KDIGO, that group has come to the conclusion that, although there is a recommendation about frequency, it is ungraded, which is basically their way of saying that there is absolutely no evidence to indicate or to support a specific recommendation.

DR. DALRYMPLE: And this Lorien.

I struggle with this similar issue. I think you have to assume that, if diagnosing and treating iron deficiency is important, then measuring iron stores is a necessary component of that. The frequency of which we should be doing it, which to deem this a true performance measure, I think is very difficult.

CO-CHAIR CROOKS: Lisa?

DR. LATTS: So, I mean this, obviously, is going to come up a whole bunch of times over the course of the next two days, and there is a lot of measures that are in this same vein.

And I guess, Karen, you said it is necessary, but not sufficient. What I would like to have seen, especially given that this is now a review of an endorsed measure, is moving on to the next step of actually assessing the outcome, the thing, as opposed to the process.

I guess it leaves us, as a Committee, sort of stuck in that we don't have that. So, do we go with nothing or do we stick with the process?

CO-CHAIR CROOKS: And just along those lines, the thing, though, has sort of become out of reach because of the new FDA ruling, too. So, in this particular metric, it may be harder to go further.

Other comments, thoughts?

DR. KASKEL: This is Rick Kaskel.

I think it becomes important in terms of resistance to ESA. Dosing is increased, and that is an algorithm now, not just one point in time that you would be measuring this. I think that was the emphasis of my scoring.

This is important. But, again, the purview of a resistance state, this would become even more important for increasing ESA dosing.

DR. PACE: One other context issue

from the recent ESRD Project, you did recommend a measure that will be endorsed about actually use of iron therapy. So, again, that is an intervention that is closer to desired outcomes, but that was specifically for pediatric patients.

CO-CHAIR CROOKS: Okay. So, I think we are ready to try to vote.

DR. PACE: Right. So, I think what we need to do is rate the evidence for the measure, which is the frequency of assessment of iron stores. Once we get the ratings on important or on this, we can then talk about whether you want to invoke the exception for expert opinion versus evidence, depending on how this comes out. So, I just want to kind of lay out what the steps are going to be.

DR. BERNS: Again, just a clarification, if I could. We are voting on the evidence that the frequency of every three months is supported as opposed to the need to measure?

DR. PACE: Right.

Okay. So, we will first rate the quantity of studies in the body of evidence.

And go ahead and start the timer.

The rating is high, moderate, low, insufficient.

MS. RICHIE: Lorien, high, moderate, low, insufficient?

DR. DALRYMPLE: For quantity?

DR. PACE: Yes.

MS. RICHIE: Yes, for quantity.

DR. DALRYMPLE: Moderate.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: Moderate.

(Whereupon, a vote was taken.)

DR. PACE: Okay. There we go.

CO-CHAIR CROOKS: The results are 13, moderate; 6, low.

DR. PACE: Okay. All right. We will move on to --

CO-CHAIR CROOKS: And 2, insufficient -- I'm sorry -- to get to 21.

DR. PACE: Yes.

CO-CHAIR CROOKS: Sorry.

DR. PACE: That's all right.

I will move on to evidence, the quality of the body of evidence. And this relates to the study design, the strength of the evidence, the directness of the evidence included here.

So, quality of the body of evidence, and the rating is high, moderate, low, insufficient. Go ahead and start.

MS. RICHIE: And Lorien, quality?

DR. DALRYMPLE: Moderate.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: Insufficient.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay. We have 2, moderate; 8, low; 9, insufficient. I'm sorry. Four, moderate.

Did I say 2? Oh, I'm sorry. I'll get this right yet. Or maybe someone else should read them.

Four, moderate; 8, low; 9, insufficient.

DR. PACE: Okay. And then, the last one would be consistency of the evidence, high, moderate, low, insufficient.

MS. RICHIE: Lorien, consistency?

DR. DALRYMPLE: Moderate.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: Insufficient.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: The results: 7 voted moderate; 4, low; 10, insufficient.

DR. PACE: Okay. So, based on our decision algorithm, basically, we have mostly lows and insufficient for the -- do you want to just show the quantity?

Quantity was fine, but then --

CO-CHAIR CROOKS: But it didn't make quality or consistency.

DR. PACE: Right. So, it basically would not pass evidence. It also did not pass opportunity for improvement.

So, let's move on to the next slide.

So, the question about evidence is we do have an exception for measures based on expert opinion, if the Committee really thinks that a measure on this topic should be considered further. We would like to see that expert opinion has been systematically assessed and fits in a guideline -- I think that is probably the case -- with agreement that benefits to patients greatly outweigh harms. Obviously, there's not -- well, I won't say that. You would be in a better position to judge benefits over harms.

So, I guess the first question, is there anyone on the Committee that wants to have a discussion about whether this measure warrants further discussion in terms of a need for this type of measure and whether you want to invoke the expert opinion exception?

CO-CHAIR CROOKS: Well, it is hard to believe that this metric causes more harm,

you know, a lot of potential harm, although, again, it is certainly possible, unintended consequences.

DR. FENVES: The other issue is -- because we are arguing about intervals, right? -- I can't foresee a study that is ever going to be done or funded comparing, say, one month, three months, six months. It just won't happen. So, it will have involve some opinion. I don't think --

DR. BERNS: I think this is one of those measures or one of these items that falls into the realm of it is probably good clinical practice, but maybe doesn't require public reporting.

MS. ANDERSON: I think it is also one of the measures there's no performance gap. When you look at the intervals and the performance gap as was related here, it is actually happening without the gap, and there is no room for improvement. I am not sure that -- clearly, there's no harm, but there

may not be any benefit to continue.

CO-CHAIR CROOKS: And along that line, one of the criteria we looked at, has the outcome improved in the last three years when this metric has been in place? I would be surprised, but do we have any information on that?

DR. PACE: Well, actually, that is where I know it would come up under validity, but maybe we should put it in the context of this discussion about whether you want to invoke expert opinion and reserve status for a performance measure.

But in the validity section, under 2b -- and maybe you want to pull that up. Okay.

So, if we look at 2b2.3, 2b2 is about validity testing. And so, what Arbor did is they looked at quintiles of performance on this measure, assessing iron stores, and looked at it in relationship to performance on the mortality measure that they have and,

basically, showed that there was an association. The two lowest performance measure quintiles had higher risk, those facilities had a higher risk of mortality. Lower performance on this assessment measure was associated with higher risk of mortality, but it was really just a difference of the lowest two quintiles. Of course, that is because performance was so high to begin with.

So, I just wanted to point that out in case that factored into any of your decisions.

CO-CHAIR CROOKS: But they didn't directly report on here's where it was three years ago and it was only at 90 percent compliance, and now it's at 99. They didn't have any data like that.

DR. FISCHER: But that may come up again, though. What if there was a measure that was endorsed and the performance gap closed, and now there isn't? So, then, do you now no longer endorse that because there is no

longer a performance gap? I mean, if we assume that that changed practice, then -- I mean I am just trying to understand properly.

Because then you have to assume -- or is there going to be some periodicity that people will assume that these will be revisited in the future? And then, if the performance gap were to evolve again, then we would come back to it? I mean I am just trying to think.

DR. PACE: And that is where we have that reserve status to continue endorsement under reserve status. So, the question is, if you think that this measure merits that, we would continue the evaluation to see if all the criteria are met for that. Okay.

CO-CHAIR CROOKS: While we are thinking about it, I guess we should have Dr. Narva introduce himself.

Does he have a disclosure-of-interest form? Okay.

DR. PACE: Go ahead.

CO-CHAIR CROOKS: All right, are you prepared?

Here's our list of things to just say briefly, so that you can join the Committee.

DR. NARVA: I'm Any Narva. I direct the National Kidney Disease Education Program at the NIH in Bethesda, Maryland. I am interested in improving care for people with CKD in the primary care setting. That is the major focus of our program.

My major non-medical interest right now is my three-year-old son.

And I don't have any disclosures.

CO-CHAIR CROOKS: Welcome, Dr. Narva.

Okay. So, we are going to vote on whether we want to kind of override our last decision and continue this anyway?

DR. PACE: Right.

CO-CHAIR CROOKS: Is that what we

are voting on?

DR. PACE: So, let's go ahead. I think there's enough discussion about it. Let's go ahead and vote on this question, which is the exception for empirical evidence based on the --

CO-CHAIR CROOKS: Well, are we voting that we are going to -- I may have confused the issue. Are we voting that the benefits outweigh the harms or are we going to vote to continue this despite the fact --

DR. PACE: Well, we first have to vote on this issue of using a measure based on expert opinion. If you agree that that's okay, then we will talk about the reserve status, right.

CO-CHAIR CROOKS: Well, the way the question is worded here is different.

DR. LATTS: Yes, will we want to continue this measure, given the lack of empirical evidence?

DR. PACE: Right. So, basically,

what we said is that there is really no empirical expert; it is expert opinion. And is the expert opinion such that the benefits to patients greatly outweigh potential harms?

CO-CHAIR CROOKS: Okay.

DR. PACE: Okay?

CO-CHAIR CROOKS: That's the question?

DR. PACE: Yes.

CO-CHAIR CROOKS: Okay.

DR. PACE: All right. So, the options are yes and no, that the benefits outweigh harms.

DR. DALRYMPLE: I'm sorry. This is Lorien. Can you just restate the question that we are to answer yes or no to?

CO-CHAIR CROOKS: Restate the question again.

DR. PACE: Okay. The question is, there is no empirical evidence and expert opinion was systematically assessed with agreement that benefits to patients greatly

outweigh potential harms. And it is judged by this Committee that the benefits to patients clearly outweigh potential harms.

MS. RICHIE: It's also Section 1c on the submission form.

DR. PACE: And the responses are yes and no. Okay.

DR. BERNS: Can you clarify the consequences of this vote, please? If we vote yes, then this --

DR. PACE: Then we will just be able to continue and discuss whether we want to consider this for reserve status. If the answer is no, we will just end here.

DR. BERNS: Okay.

DR. PACE: All right. Sorry. Okay.

CO-CHAIR CROOKS: Okay. Go ahead and start the vote.

MS. RICHIE: Okay. So, Lorien, yes or no?

DR. DALRYMPLE: I'm going to say

no.

MS. RICHIE: And Kristine, yes or no?

CO-CHAIR SCHONDER: Yes.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

DR. PACE: And now we need 22 votes.

CO-CHAIR CROOKS: Yes, 22 votes, right.

DR. PACE: Andy, are you voting on this? Okay.

CO-CHAIR CROOKS: Okay. Stop there.

DR. PACE: All right, go ahead, stop it.

CO-CHAIR CROOKS: Results: 16, yes; 5, no.

DR. PACE: Okay, so let's move on to the next slide, Tenee.

So, basically, we still have a measure that did not pass performance gap or

performance, opportunity for improvement. So, let's go on to the next slide.

So, it came up, do we want to consider this for potential for reserve status? If so, it means we will continue to assess reliability and validity. If not, if you think it is still not going to be -- you know, because there is no performance gap, but it is not going to be that useful for continuing for potential for reserve status. And so, well, the question is whether this even meets our criteria for reserve status.

It is not proximal to desired outcome. There is no strong direct evidence. It just expert opinion. It is obviously related. So, I don't know. What do you think?

DR. BURSTIN: I think it is a close call. I think it is really up to this group really who knows the evidence best to make that assessment. I mean, to me, in general, it seems somewhat analogous to the fact that

we look at the levels of A1C control in diabetics. There is a measure that looks at did you measure A1C levels. That is sort of, I think, moving its way out of fashion and moving more towards just looking at the outcome.

I think that is the decision the group needs to make. Is there still value in the assessment measure when you can also look at the intermediate outcome?

DR. PACE: So, we are still at the point of it is important to do in clinical practice. Is it something that you want to consider for reserve status?

DR. BURSTIN: It is really reporting of the assessment, right? So, it is still not getting at the --

DR. PACE: Okay.

CO-CHAIR CROOKS: Yes?

MR. MESSANA: Joe Messana from UM KECC.

I just wanted to make sure, in the

performance gap discussion there was some difference between the results for hemodialysis and peritoneal dialysis patients, 97 versus 86 percent, and whether that would influence a discussion of whether there was a performance gap for a subset of patients.

CO-CHAIR CROOKS: Okay. So, the vote, then, is whether we are going to grant this reserve status.

DR. PACE: Well, whether we will continue evaluating.

CO-CHAIR CROOKS: Continue evaluating.

DR. PACE: That won't be decided until the end. Otherwise, it will stop here.

CO-CHAIR CROOKS: The potential for reserve status?

DR. PACE: Right.

DR. LATTS: And again -- I'm sorry -- if it is in reserve status, it means that it is not in active use but it is out there for the future in case the performance gap

widens. But how do we know if the performance gaps widens if it is not in active use?

DR. BURSTIN: So, the idea would be that there would be some ongoing surveillance, but not public reporting, periodically to make sure performance doesn't fall down. But it may not rise to the level of what we think is the importance of other measures with a known gap.

Again, this is relatively new for us. A handful of measures have been put into reserve status to date as part of the Cardiovascular Committee. And again, very similar sorts of discussions.

Really, the idea would be that the measure would not have to go through a full endorsement later when it comes up for maintenance in three years. There could be a discussion that says, actually, the background surveillance, they have got this CROWNWeb anyway, would suggest, actually, there has been a decrement of performance. Maybe we

should move it back up to active use.

DR. PACE: Just in context, though, I think those cardiovascular measures were mainly interventions, not assessment measures.

DR. BURSTIN: Yes, process measures.

DR. PACE: Right. I said they were intervention measures. So, what is an example of one of them?

DR. BURSTIN: Aspirin use, for example, in the context of AMI, it is hard to walk into any emergency department in America without an aspirin in your mouth. So, things like that --

DR. PACE: Right.

DR. BURSTIN: -- are what the Cardiovascular Committee considered.

I will tell you that the appetite for assessment measures is one that always gets complicated when they go through the process.

DR. PACE: Further discussion?

(No response.)

Okay. So, what we are voting on now is whether you want to continue looking at this measure for potential reserve status, yes or no.

Okay, start the timer.

MS. RICHIE: And Lorien, yes or no, reserve status?

DR. DALRYMPLE: No.

MS. RICHIE: And Kristine, yes or no?

CO-CHAIR SCHONDER: Yes.

(Whereupon, a vote was taken.)

DR. PACE: Okay, one more person.

CO-CHAIR CROOKS: Wow, that last vote.

Okay, we have 10 yes and 11 no.

DR. PACE: Okay. So, we will actually stop on this measure here. It would not go forward to be recommended.

When something like this is a fairly close vote, we kind of highlight that

in the draft report. Then, we will see what kind of comments we get. You would have the potential to reconsider that, but at this point our votes are based on majority vote. But we will definitely note that in the draft report.

Okay. So, I know it feels like we are a little behind time, but part of this is kind of getting down the process. I think we are doing okay.

Shall we move on to the next measure?

CO-CHAIR CROOKS: 1660.

DR. PACE: Okay. So, that is Rick Kaskel.

DR. KASKEL: Ready?

CO-CHAIR CROOKS: Go ahead.

DR. KASKEL: A brief description of this measure is it is the percentage of calendar months within a 12-month period during which patients age 18 years and older with a diagnosis of ESRD who are receiving

hemodialysis or peritoneal dialysis have a hemoglobin level less than 10 grams per deciliter.

The numerator is the calendar months during patients having a hemoglobin level less than 10 for the last hemoglobin quarter for each calendar month, and the denominator is the calendar months during which patients age 18 and older are receiving dialysis.

A series of exclusions are listed there. It is an outcome measure, and the source is from administrative claims, electronic clinical data, health record registry and paper. And it is a clinician group, individual, or team. And the measure is not paired, nor is it a composite.

DR. PACE: Okay. So, we will start with if you would just summarize what the preliminary was for impact?

Tenee, are you ready to start?

So, if you want to just do that

first, and then we will vote on that, and then move on to the next one.

DR. KASKEL: It looks like the results were split between high and moderate, well, actually, favoring high more than moderate. Importance was, again --

DR. PACE: Okay. We will do them one at a time.

So, are you ready, Tenee?

Any comments or issues about impact for this measure?

DR. DALRYMPLE: Hi, Karen. This is Lorien again.

I was actually hoping to ask the primary reviewers their thought on impact. When I reviewed the measure, I thought sufficient data was presented on why less than 10 versus less -- an important performance measure, for example.

DR. PACE: And I think we will hold that for the evidence discussion related to what perhaps the threshold is to be. But this

is more general about the topic area, the impact on patients. We will get to the specifics regarding -- you know, I think that will come up more under evidence.

DR. DALRYMPLE: Okay. Thank you.

DR. PACE: Okay. So, let's go ahead and vote on impact for this measure. This is 1660.

MS. RICHIE: And Lorien, high, moderate, low, insufficient?

DR. DALRYMPLE: Moderate.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: High.

(Whereupon, a vote was taken.)

DR. PACE: Okay. Tenee?

CO-CHAIR CROOKS: Twelve voted high; 9, moderate; 1, insufficient.

DR. PACE: Okay. So, let's move on to opportunity for improvement or performance gap regarding this measure.

Rick?

DR. KASKEL: It looks like the

reviewers were split between moderate and high.

DR. PACE: Okay. Any particular comments about performance gap on this that any of the reviewers or Steering Committee want to point out?

DR. DALRYMPLE: Sorry, this is Lorien again.

I am just hoping to get clarification. What percentage of patients were less than 10 as opposed to not between 10 and 12? I wasn't sure what was meant by optimal care.

DR. PACE: Lorien, you're breaking up a little bit.

DR. DALRYMPLE: Oh, sorry, Karen. Let me try that again.

I was just wondering if maybe someone else on the Steering Committee could clarify if they were able to delineate what percentage of patients were less than 10 as opposed to not between 10 and 12. I wasn't

sure what was meant by optimal care on page 4, if that was showing the percent of patients who were falling less than 10, or what was meant by that data.

DR. PACE: Okay.

CO-CHAIR CROOKS: Yes, I think what she is asking, and I agree, the way the data is presented, it is 10 percentile, but it doesn't tell me what percentage of patients were below 10. Is that the 36.5 percent?

DR. PACE: So, we are looking at page 4 of the submission for 1660?

CO-CHAIR CROOKS: Right, 1b.2

DR. PACE: Okay. Let's ask the measure developer.

MS. CHRISTENSEN: Yes, I am Keri Christensen, AMA PCPI.

Just to clarify that, that is 36.51 percent of the patients didn't meet the measure for that year. Then, the percentiles are, if you were a provider at the tenth percentile, you would have had 10.42 percent

of your patients meet the measure.

DR. BERNS: If I could, just a comment. We are in a period of transition here where the prior assessment of this as a performance measure was looking at those who were below 10 as being bad, and now we are in a period of time where, at least as far as the FDA is concerned, below 10 maybe isn't so bad.

So, I think it is impossible to assess the performance gap or we have to look at any data on a performance gap through a very different set of eyeglasses than we were before because that is no longer considered necessarily a bad thing. Whereas, I think, particularly when this was done in 2008, but even a year ago or two years ago, the general consensus was that a hemoglobin below 10 was bad. I am not sure we are at that same place anymore, at least as a universal brush with which we are supposed to be painting all of our patients.

DR. NALLY: And specific to this

measure, I am sure we will have more discussion in the adult measures, which I understand have been withdrawn. Is there a specific --

DR. PACE: Let me just clarify. This is an adult measure --

DR. NALLY: Yes, I understand.

DR. PACE: -- for physician performance.

DR. NALLY: I misspoke.

So, can we broaden out this discussion somewhat in terms of the implications of this? Are we going to do that during this evidence phase?

DR. PACE: Yes, I think we will get to that just momentarily when we talk about the evidence, and definitely need to address that because it affects multiple measures here. And as I mentioned earlier, it is the reason that CMS withdrew their pediatric and their adult, which were facility-level measures. These measures that we are looking

at now are being presented as physician performance measures.

DR. NALLY: Thank you.

DR. PACE: Okay. So, the comment was made that it is hard to interpret performance gap, not knowing what the target should be or what the acceptable target.

CO-CHAIR CROOKS: Okay. So, before we vote, any other comments before we vote on performance gap?

DR. LATTS: I guess the sort of takeaway for me on that is that we may not know whether high is better or low is better. Or we may not know what the right performance is on this measure. But it actually suggests there is definitely a gap, and important to measure for that reason because we don't know what performance is.

DR. KLIGER: And I guess the confusing thing is that it says less than optimal. If we don't know optimal, it is hard to judge these data at all. That's the

problem.

DR. FISCHER: But I thought it is an "or" statement, right? Either it is suboptimal or there is variation, right? I just want to make sure I understand because my understanding was that, if there is variation, we don't have to worry about what is optimal and not optimal. If there is variation, then that means that it could be potentially a performance gap? Or do I have a misunderstanding?

DR. PACE: No, you're right. The performance opportunity for improvement, I mean the classical QI perspective is variability in performance or if there is overall suboptimal. And we will definitely address the evidence question.

But I think the comments are important. I mean one way you could address those is that maybe we don't know at this point because we don't know what the optimal is.

But, anyway, you're right in terms of what our criteria are about.

CO-CHAIR CROOKS: Okay. Other comments?

(No response.)

Let's vote.

DR. PACE: All right. So, this is performance gap, high, moderate, low, insufficient.

Go ahead and start.

MS. RICHIE: Lorien, high, moderate, low, insufficient?

DR. DALRYMPLE: Moderate.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: High.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: We have 6 who voted high, 8 who voted moderate, and 8 who voted insufficient, I presume because we don't know what the performance should be.

How do we write that?

DR. PACE: Well, let's continue on with the evidence discussion. Then, I think it will become more clear what you all want to do.

CO-CHAIR CROOKS: Okay.

DR. LATTS: I just have got to say I am really glad you guys started with an easy one to get us going.

(Laughter.)

DR. PACE: All right. So, Rick, do you want to talk about the evidence?

And we will talk about quantity, quality, and consistency, but I think the big issue that everyone is aware of is the recent FDA announcements and how that pulls in here.

DR. KASKEL: Would it be helpful just to review briefly some bullet points on the FDA? I have it here.

DR. PACE: Yes. Definitely.

DR. KASKEL: Okay. Do you have it?

DR. PACE: Okay.

DR. KASKEL: Is it in the handout

at all as a slide?

DR. PACE: Alan?

DR. KLIGER: I just want to comment it is not just the FDA.

DR. PACE: Right.

DR. KLIGER: I don't think it is just simply the FDA's announcement, but the body of data that we have come to understand in the last year.

DR. PACE: Good point.

DR. KASKEL: So, should we read it, so we all know it?

DR. PACE: If you have bullet points, why don't you just briefly highlight --

DR. KASKEL: Okay. So, the FDA drug safety communication modified dosing recommendations to improve the safe use of ESA and CKD, they made these recommendations because of data showing increased risk of cardiovascular events with ESAs in this patient population. And there was a box

warning that basically said, in controlled trials with CKD patients, patients experienced greater risk for death, serious adverse cardiovascular reactions, and stroke, when administered ESAs to target hemoglobin levels greater than 11. No trial has identified a hemoglobin target level or ESA dose or dosing strategy that does not increase these risks.

And in patients with CKD, consider starting ESA treatment when the hemoglobin is less than 10 grams percent. This advice does not recommend that the goal is to achieve a hemoglobin of 10 or greater. Individual dosing is recommended.

So, basically, there is nothing mentioned about the lower target here.

DR. SOMERS: I actually wanted to ask a question to the measure developers. I don't know whether this is the question or not.

CO-CHAIR CROOKS: Please go ahead.

DR. SOMERS: But I wanted to

understand why, given that the facility measure has been withdrawn, if this a physician measure, if you had some comments as to why you thought it would be important to go forward with the physician-level measure in that setting?

MR. JONES: Thanks for the opportunity.

First of all, remember, the facility-level measure was a payment measure. When the KWIP changed from no longer having the minus 10, it wasn't part of that program. And therefore, it was removed because it was a payment, I believe payment method. Whereas, we are talking about a physician-level measure that is for public reporting.

The Work Group within PCPI evaluated the current situation after the FDA announcement and felt that the measure should stay in place.

The reason for that is a few-fold. One is the fact that it was not saying that a

certain percent of patients should be above 10. It was really looking at it as a patient safety issue. As the hemoglobin falls down, as you know, an inflection point of actually 11, the number of transfusions increases. We saw that we would not have some way of measuring the safety effect of increasing transfusion in a vulnerable population.

And with the changing pattern on the use of ESAs, trying to look and see whether there is going to be a normative use of this drug, having a higher and a lower measure we felt was also important to develop future patterns.

So, we saw it as a safety issue of how are we going to track the issue of patients getting ESAs, and, As you all know, the increase incidence of sensitization and increase incidence of not getting transplants.

If I could just make one other comment regarding the FDA, remember a year ago a number of folks, some in this room, experts

in this areas presented to the FDA Advisory Panel and presented some of the same data you are talking about and know today. That experts panel's opinion was that at that time the label should not change, that it should stay the same, with the available data that was there, including after the TREAT and the other data.

The FDA elected to ignore the advice of the Advisory Panel. So, I think we have to keep that in mind, that the experts who testified and the Advisory Panel itself advised FDA not to change.

And Barbara Fivush is also another person --

CO-CHAIR CROOKS: Okay. Kathleen?

MS. LEBEAU: Thank you.

I would really like to piggyback on that because, while I really understand this is an evidence-based process, this is exactly what I am hearing in the patient community, that without that failsafe, every patient I

talk to is very concerned who understands this that their hemoglobins, the average hemoglobin is going to drop. And what that means, all the resultant complications, quality-of-life issues, risk of transfusion, potential for transplant, takes a big toll on the patients.

So, I think everything that Dr. Jones just said is echoed multiple-fold times within the patients community.

CO-CHAIR CROOKS: Thank you.

Who else? Barbara?

MS. FIVUSH: So, many of you I am sitting on a different side of the table now than I did last time. I would support Ed's comments and remind everybody, and you are going to see this as a pediatric measure -- I know you have all seen this in your packets -- that we have this distinction about age 18, below which you are a pediatric patient, above which you are an adult. But there are many young adults in this population that are going to fall into this measure category that are

really still pediatric, in our mind as pediatric nephrologists, they will need transplants. If they require multiple transfusions, this is going to impact their life long-term.

We have data on quality of life in these young adults who are still going to school, who are still try to get to vigorously exercise.

So, although we support and we understand that this is going in as an adult measure, I would just like you to consider this also as an older pediatric measure and the impact of really young adults in this measure.

CO-CHAIR CROOKS: I would like to just piggyback off that and kind of mention the issue of sensitizing patients who might be getting a transplant.

If this turns into like you are looking at Physician A, B, and C, and Physician A has a population with low, a lot

of patients below 10, but that might be the doc who is trying to avoid transfusions so that his patients can get transplanted or have a higher chance of being transplanted successfully.

So, what bothers me about this is the implication that it becomes a good and bad thing. You are looking at each doctor, and because I have 30 percent below 10 and the other doctor is 10 percent below 10, does that mean I am bad compared to the physician with 10 percent? I don't think you can conclude that.

DR. LATTS: Well, and I guess the question is, given all the controversy around this, is it important to measure because of the controversy as opposed to overlaying a doctor who performs X or Y is bad? And I don't know the answer to that. I am just putting it out there.

Because this is such an unknown now, is that another reason why it is

important to measure without the overlay of what good or bad performance is?

DR. NALLY: I mean, Ed, I appreciate and understand everything you say, and probably believe all of that in my own practice. But the issue of controversy here is whether we send this mixed message where we have FDA and CMS trying to release one message again, but this is where things stand, and then have a variation on the theme of that. That is the concern I have about the mixed message. And theoretically, FDA and CMS made a decision based upon the best available data with a group of people impacting it.

I would be curious to hear Jeff's opinion because he was before that Committee you mentioned, and as far as I know, is probably he and Alan may be the two people in this room with the most insight into this.

DR. BERNS: I wasn't at the FDA hearing. I actually spoke yesterday with somebody from the FDA who had contacted me. I

think that the FDA's emphasis is switching, although it only merits one line in their black box warning and their new labeling, which is individualization of anemia management.

I mean, ideally, the performance measure that we would have is a percent of physicians or patients who have appropriate hemoglobin levels for them. How we are going to accomplish that I have no clue.

But I think that, one, is that the TREAT acquire, create normal and adequate data, were obtained on a very specific patient population for which many of our patients fit, but many of our patients don't. I think we have to take the responsibility of making sure that each patient is treated appropriately, have the appropriate hemoglobin, the appropriate mix of iron and ESA therapy, avoidance of transfusion or not, depending upon the individual circumstances.

And I think this is going to

create, this particular measure will create huge amounts of confusion because we don't know whether this is good or bad because it is good or bad depending upon the patient. At the physician level, it is good or bad depending upon the mix of patients that that practitioner is taking care of, whether it is a largely pediatric population, whether it is a young, healthy population, whether it is a nursing home population. And the best percent of patients below 10 may range between zero and 100 percent, depending upon what type of patient population a practitioner is taking care of.

DR. KLIGER: At Joe's invitation to say a word, the data that we have looked at all deals with the higher hemoglobin levels and use of ESAs to achieve higher hemoglobin levels. The data doesn't address at all lower hemoglobin levels, and it is silent about hemoglobins around 11.

So, my own personal take on this is

that we all know that hemoglobins of less than 6-7 are really bad for people. I don't think anyone would doubt that. But the data around where this measure is is absent. I surely think that measurement is important. I think Lisa makes a very, very good point in an area where we are unclear. But the question before us is to endorse or not endorse a measure that talks about lower than a level for which there are just no data at all in the range of 10.

DR. LATTS: And, Alan, just to follow up on that, is there anything that the developers could do to this measure that would get to what that critical gap is that we need to know? I mean, you know, if they dropped it to 7 or 8, would that be something meaningful or would there be an above or a below threshold?

DR. KLIGER: Yes, again, I think that is the right question. I think that, were we able to review data at some level where we could show the distal effects of

health, that would be very useful for us, for me.

CO-CHAIR CROOKS: Ruben?

DR. VELEZ: I'm somewhat confused because we are using 10 in this measure, and that is exactly that magic number that the FDA came with, although this measure doesn't look at ESA, but they said 10, below 10, you may be prone to use ESAs, if needed. But they brought this up.

So, I do not see a big issue with what the FDA said and this measure, in particular. Now, yes, it would be measuring patients with hemoglobin of less than 10, and that's it, without saying good, bad, ugly, or the rest. But I think it does go hand-in-hand with what the FDA came out with.

CO-CHAIR CROOKS: Yes, Roberta?

MS. WAGER: Dr. Berns, I have a quick question for you. When you talk about appropriately treating the patient, what is your definition of appropriate? And does that

include your feedback from the patient on how lousy they feel at hemoglobin 9, can't work, and how great they feel between the 10, 11, and the 12?

DR. BERNS: Yes, absolutely, and this is the discussion I had with the gentleman from the FDA yesterday, was that I have lots of patients who I think will be inadequately treated if the results of TREAT and the other clinical trials influence rules or guidelines about how those patients should be treated.

And this is a problem with clinical practice guidelines. It is a problem with performance measures, is that they apply to everybody.

And we have had this vigorous debate on the KDIGO panel. I am not sure it has been entirely resolved because there is a great deal of sort of a sense of urgency, in an effort to protect patients, to get away from this sort of push to ever increase ESA

doses and iron and hemoglobin levels. But, again, I think as we all recognize, the risk is that there will be patients who are ill-served by having hemoglobins of 10.

Just sorting that out and then converting that to a manageable performance measure is the problem. That is the dilemma.

CO-CHAIR CROOKS: Andrew?

DR. FENVES: I just want to make one comment. I completely agree with Jeff.

And to comment on variability in practices, if you have a practice with older patients with high cardiovascular burden, with symptomatic anemia, for example, or it could be symptomatic at very different levels, then you are going to be very aggressive depending, maybe even at higher levels, as opposed to a very young population with good hearts, let's say, and much less burden; you might allow -- I mean the literature in hematology would say you can let hemoglobin fall to fairly low levels.

CO-CHAIR CROOKS: Kathleen?

MS. LEBEAU: Thank you.

I did testify at that FDA hearing a year ago. Lisa and Alan, that is exactly, when they polled the Advisory Committee members, that is exactly why they came to the decision they did. There were simply no studies. And that was their recommendation, was that more studies should be done.

Now the FDA, going against its Advisory Committee, when that hasn't been resolved, and what that means for our decision today, as Lisa says, I'm putting that out there.

DR. BERNS: I'm sorry, one comment about the FDA, and maybe others can correct me if I am wrong. Their mission is to minimize harm, not maximize benefit, I guess. So, I think we do need to keep that in mind, that they are responding to risk of harm that was made apparent by actually a relatively small number of clinical trials, but a relatively

large number of patients. Very different than thinking about individual patient potential benefit.

CO-CHAIR CROOKS: Jerry?

DR. JACKSON: There is a tendency for a measure to be looked at as anyone falling into the percentage of less than 10 being high as underperforming in some way. That has a tendency at the network level of being looked at as a problem, even though this nuance of how it should be interpreted may be understood at this Committee, but not so much at the network.

So, the point being that -- and this is a theoretical -- but if the number of 10 is chosen, and you are in that category of having a high percentage of people under 10, you are going to tend to push the ESAs so that you don't let your overall population fall down to a certain percentage under 10.

And it is very difficult, as we all know, to tightly regulate ESAs. I think it is

just going to shift the curve back up to a higher percentage of people being over 11, if we are trying to avoid being under 10.

I am conflicted on this because I still think, for all the reasons that Dr. Jones mentioned and others, that this is a valid measure, but it does have the unintended consequence of pushing the average hemoglobins within a clinic up, so that you are going to have a certain percentage that will then shoot over the top and get too high.

CO-CHAIR CROOKS: Right, and in that way, it could be unintended harm. In other words, it is going to encourage treating more patients with ESA than you would, if you are being evaluated on that criteria.

We have the developer wanting to make a comment. Anybody else before we go there?

(No response.)

Okay, Diedra?

MS. JOSEPH: Hi. Deidra Joseph

with the AMA PCPI.

We just wanted to point out that there are several denominator exclusions for the measure. We have documentation of medical reasons for patient having a hemoglobin level less than 10 grams per deciliter. And we have listed examples here, including patients who have non-renal etiologies of anemia, including sickle cell anemia or other hemoglobinopathies, multiple myeloma, primary bone marrow disease, anemia related to chemotherapy for a diagnosis of malignancy, and other medical reasons.

CO-CHAIR CROOKS: Okay. So, Karen, how do we proceed?

DR. PACE: All right. So, I guess what we will do is vote on the quantity, quality, and consistency of the body of evidence, and then we will see where we are at.

Alan, did you have a question? Okay.

DR. BERNS: Just a clarification?

DR. PACE: Yes.

DR. BERNS: On this item, it is the quantity of studies addressing this specific hemoglobin level --

DR. PACE: Right.

DR. BERNS: -- not whether or not anemia is a good thing or a bad thing?

DR. PACE: Right.

CO-CHAIR SCHONDER: Hi. This is Kristine.

I wanted to make a comment on the body of evidence. Because we have heard a lot of discussions both for and against the actual hemoglobin target. But one of the things that I was trying to do, as I was reviewing this particular measure, is look at the evidence that they were presenting specific to the consequences of a hemoglobin level less than 10, as opposed to being within a goal range. I think the body of evidence that they are presenting really is pointing to having a

hemoglobin target within a particular range.

CO-CHAIR CROOKS: Okay.

DR. PACE: Okay. So, let's vote on quantity, and the options are high, moderate, low, insufficient.

MS. RICHIE: And Lorien, quantity?

DR. DALRYMPLE: Insufficient.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: Insufficient.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, 1, high; 2, moderate; 4, low; 15, insufficient.

DR. PACE: Okay. So, let's move on to the next one, which is quality of the body of evidence. Again, the options are high, moderate, low, insufficient.

And go ahead, Tenee.

MS. RICHIE: And Lorien, quality?

DR. DALRYMPLE: Insufficient.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: Insufficient.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Andrew, are you voting now?

DR. PACE: One more vote.

Okay, everybody has voted.

CO-CHAIR CROOKS: I guess that's all the votes we're going to get.

DR. PACE: Go ahead.

CO-CHAIR CROOKS: Okay. Three votes for moderate; 3 for low; 15 for insufficient.

DR. PACE: Okay. And then, finally, consistency of the body of evidence. And again, the options are high, moderate, low, insufficient.

And go ahead and start the timer.

MS. RICHIE: And Lorien, consistency?

DR. DALRYMPLE: Insufficient.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: Insufficient.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Is everyone voting? Anybody not voting? We're missing one for some reason.

DR. PACE: Okay.

CO-CHAIR CROOKS: Someone went to the restroom. Okay, let's go ahead.

The results are 2, moderate; 1, low, and 17 insufficient.

DR. PACE: Okay. So, let's just sum up. It obviously did not pass evidence. And where were we on opportunity for improvement, 1b?

Okay. So, basically, this measure would not pass the importance to measure and report criterion.

Any further comment on that?

(No response.)

Okay. So, let's go on to --

CO-CHAIR CROOKS: Alan?

DR. PACE: Oh, Alan?

DR. KLIGER: Is there a way for us, though, to register -- I want to go back to

Lisa's comment before. Because it doesn't pass, but if we take a consensus here of the importance to be measuring hemoglobin levels in our dialysis patients, I think you would have 100 percent agreement to that.

So, this specific measure doesn't make it, but it would be important somehow for us to register our concern that hemoglobin is an important measurable intermediate outcome in our patients.

DR. PACE: So, this gets to, then, having a performance measure on just whether it is being assessed. Is that what you are suggesting? And this may be a good example of when that is needed, based on the fact that the evidence doesn't support a specific target.

I guess the question would be, is there anything else that could be measured that is more proximal to the desired outcome? Is there any treatment approach? And I know that that is individualized as well. So, I

just wanted to lay out there to see if there was evidence that would support a performance measure that was more proximal to the outcome.

CO-CHAIR CROOKS: I don't think so. I think that is about as close as we can come to the outcome. The outcome is excellent anemia management for each individual patient. And because it varies so much, maybe the closest we can get is that. At least the hemoglobin level, the iron levels are being looked at. Okay.

DR. PACE: Okay. All right.

CO-CHAIR CROOKS: So, according to my agenda, lunch is supposed to be ready at 12:30. So, we could actually -- although we need to have a little time for comment, but should we go on to the next?

DR. PACE: Well, I wonder, the next two measures are in, let's see, we have 1666, which is about greater than 12. Should we maybe talk about -- 1667 is the same measure, but the pediatric version, less than 10.

So, let me just ask, and maybe we will just do -- do people feel like you would vote the same way on the pediatric measure that you just voted on the --

DR. SOMERS: I mean there are specific pediatric data that I think needs to be discussed.

DR. PACE: Okay.

DR. SOMERS: And what has been discussed isn't germane to that.

DR. PACE: Okay. All right. Then, we won't go there. We will continue through the measures then.

Do you want to try to do 1667? Or do you want to do --

CO-CHAIR CROOKS: I think we should try to do one more before lunch --

DR. PACE: Okay.

CO-CHAIR CROOKS: -- at least get started on it.

DR. PACE: All right. Well, then, let's go ahead according to schedule, 1666.

CO-CHAIR CROOKS: You know, just as a rule, we don't call on the developers unless we have a question for you. So, you can't just raise your hand and expect to be called. You can raise your hand and we might, but we might not.

(Laughter.)

So, I don't think we have any questions for you right now.

DR. PACE: Right, and when we have public comment, if you haven't had a chance, you can do it during the comment period as well.

MS. JOSEPH: We just had a question about the format, kind of the benefits versus harms discussion. We didn't know if that would apply to every measure or not.

CO-CHAIR CROOKS: No.

DR. PACE: No.

CO-CHAIR CROOKS: Only under certain conditions that are outlined in the evidence table.

MS. JOSEPH: Thank you.

CO-CHAIR CROOKS: Okay. So, we are going to go to 1667.

DR. PACE: We'll go ahead in order, 1666.

CO-CHAIR CROOKS: All right. Oh, 1666. I'm sorry.

DR. PACE: Okay. So, that's Ruben.

DR. VELEZ: Now that we have answered all the questions we need to answer, this should be easy. It should be very easy.

(Laughter.)

1666 is really patients on ESA with a hemoglobin level of over 12. Essentially, it is a percentage calendar month on a year, 12 months. And hemoglobin is measured -- this is an adult. So, patients over 18 years old with a diagnosis of CKD and ESRD, so CKD Stage 4 and 5, and ESRD, both on hemo and peritoneal dialysis, who are also receiving ESAs and had a hemoglobin of over 12.

Now they asked that this hemoglobin

is the last hemoglobin done in that month. Again, this is a calendar month. The numerator, as we said, people with hemoglobin over 12, with the denominator being people on ESAs, adults over 18 years old, with CKD 4, 5, and ESRD.

It is an outcome measure, and we could go directly to the impact, if that is okay.

DR. PACE: Yes.

DR. VELEZ: On the impact, we have four, of the Work Group, we have three highs and two moderate.

Any comments? Anything from the Work Group members?

(No response.)

DR. PACE: Anyone else want to make any comments? Otherwise, we can go to vote.

All right. So, we are voting on impact for Measure 1666, as described, and options, high, moderate, low, insufficient.

And go ahead, Tenee, start the

timer.

MS. RICHIE: And Lorien, impact? Lorien, are you still there?

DR. DALRYMPLE: Oh, yes, I'm sorry. High.

MS. RICHIE: And Kristine, impact?

CO-CHAIR SCHONDER: High.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Do you have 22? Okay.

The results are 16 votes for high; 5, moderate; 1, low.

DR. VELEZ: It must mean that we're hungry.

(Laughter.)

Okay. On the opportunity of improvement, again, the members, we had one high, two M's, moderate, and one high.

CO-CHAIR SCHONDER: This is Kristine.

On insufficient evidence, I just had a question of clarification. In the data

that was presented, they are talking about 63.5 percent of patients did not receive optimal care. I am just double-checking to make sure, does optimal care mean the measure specifications itself?

CO-CHAIR CROOKS: That's a good question. For the measure developer, that does mean, that means -- let me just rephrase it then.

Sixty-three point five percent of patients in a calendar year had a hemoglobin greater than 12 at least one time?

MS. CHRISTENSEN: It is 63.5 percent of patients did not meet the measure.

CO-CHAIR CROOKS: Did not have a hemoglobin greater than 12? So, then, 37 percent of patients during the calendar year had one or more hemoglobins 12 or greater, or greater than 12?

MS. CHRISTENSEN: There's a lot of negatives there, isn't there?

CO-CHAIR CROOKS: Well, we need to

know what -- it will help us judge if there is a performance gap or not.

MS. CHRISTENSEN: A patient meeting the measure would be a patient who has a hemoglobin level greater than 12. A patient not meeting the measure, which would be 63.5 percent of patients, did not have a hemoglobin level greater than 12 in patient months.

CO-CHAIR CROOKS: You're defining optimal care as having a hemoglobin --

MS. CHRISTENSEN: Optimal care is meeting the measure.

CO-CHAIR CROOKS: Well, but optimal care is actually not meeting the measure. That is the problem, one problem with the way this is written.

MS. CHRISTENSEN: Yes, it is poor wording.

CO-CHAIR CROOKS: Okay. We would view it as a safety measure, and if you're high, that's negative.

So, anyway, 37.5 percent of

patients, therefore, did not have a hemoglobin greater than 12. Is it once or more?

CO-CHAIR SCHONDER: I think I heard that the other way around.

DR. BERNS: But, again, the measure is months. Does this data refer to months? And does this data refer to CKD 4 and 5, not on dialysis?

CO-CHAIR CROOKS: This is just --

DR. VELEZ: This measure is a combination of two measures because it uses CKD 4, 5, and ESRD. In the past, we had just an ESRD measure. So, this one combines both groups.

DR. BERNS: The measure does, but I am not sure their performance gap data includes non-dialysis patients.

DR. VELEZ: Correct. You're completely correct.

So, if we may summarize again, can we summarize again what really that 63.5 percent that did not receive optimal care

means?

MS. CHRISTENSEN: Yes, cross out "did not receive optimal care" and put in the words "63.5 percent of patients did not meet the measure." So, 63.5 percent of patients did not have hemoglobin level greater than or equal to 12.

CO-CHAIR CROOKS: So, 37 percent above is still a significant amount, although we all know that in the variability of actual clinical practice patients, hemoglobins wander around a lot, and a lot of patients will just get over 12 once or twice during the year.

So, again, you wanted to use the words "optimal care". In other words, is it really bad if somebody exceeds this measure? Is that something that is implied by passing this?

PARTICIPANT: That is not the question here. The question here is precisely whether there is a performance gap for this measure. That's all.

CO-CHAIR CROOKS: I hear you, yes. Okay. So, let's vote.

DR. PACE: Okay. All right. Ready to vote. Performance gap on this measure, high, moderate, low, insufficient.

Go ahead, start.

MS. RICHIE: And Lorien, performance gap?

DR. DALRYMPLE: Moderate.

MS. RICHIE: I'm sorry, moderate?

DR. DALRYMPLE: Yes.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: High.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay. The results are 7, high; 12, moderate; 3 insufficient.

So, then, we can go on to -- this is not --

DR. VELEZ: This is an outcome measure, yes.

CO-CHAIR CROOKS: This is an health

outcome?

DR. VELEZ: Yes.

DR. PACE: Well, this is actually an intermediate clinical outcome. So, we will skip this question --

DR. VELEZ: Okay.

DR. PACE: -- and let's talk about the evidence.

CO-CHAIR CROOKS: Okay, Ruben, yes.

DR. VELEZ: Now, in the evidence, I mean, it is clear that the owners of the measure did mention, like we have said in many other words, there is lack of information to support a specific hemoglobin cutoff value. So, it renders this, I mean, a lot of this evidence, like we have discussed, comes from clinical practice guidelines and some of the RCTs that they mention in the information that was given.

So, I mean, we can be talking about this for a while. So, I will open for comments.

DR. NALLY: I guess one concern that I had in recent FDA/CMS releases was that there may be harm when the hemoglobin is high with ESAs, particularly at high dose. The question is the threshold for that high, whether it is 11, 12, or 13.

More recently, 11 and higher is being brought into question. We have opted to use the 12 number. So, that threshold would be the specific concern I had with using that number.

DR. KLIGER: I guess my comment is that I think we need to consider this as a safety measure. We know that achieving -- sorry -- we know that targeting hemoglobins in the high range, 13, and achieving levels in the 12s somewhere, has a higher incidence of harm than we mostly would be comfortable with.

So, my interpretation is that this is a measure that is monitoring a safety signal, and we don't have an achieved hemoglobin level, evidence for a level of

achieved hemoglobin that will give us an adequate cutoff.

So, I think Joe's comments are well-taken. I just would point out that there is a difference between complete absence of evidence at the low end, which is what we talked about before, and a safety signal at the high end, and the Committee needs to consider that as we make our decisions.

DR. BERNS: I think the other comment to make, again, although I am not sure how this is going to translate in practice, this is percentage of calendar months that a patient has a hemoglobin above 12. So, if a patient has a hemoglobin above 12 in one calendar month, then that would only count as whatever that is, 9 percent or 8 percent for that. So, it is maybe a better way of looking at some of these things than just saying the percent of patients who are above 12, because that is not really an important number.

CO-CHAIR CROOKS: I would like to

hear a little bit more about what you are saying, or maybe I couldn't hear you very well. But you are saying this is not just a percentage of patients, but it is the percentage of time in a year that a given patient --

DR. BERNS: Yes, if I interpret this correctly, it is the percentage of calendar months during which a patient has a hemoglobin above 12. So that, if a patient has, I guess looking at the way we used to do this is the percentage of patients who have a hemoglobin above 12; you just add them all up, and if they have a hemoglobin above 12 one month out of the entire year, during that month they get counted the same way as a patient who has a hemoglobin above 12 for eight months out of the year.

This way, in this measure, if I understand it correctly, it is sort of scales that, so that the patient who is above 12 for six months a year is recorded differently than

a patient who has a hemoglobin above 12 just one month out of the year.

And maybe the measure developer can correct me if I'm wrong, but that is how I interpret this measure.

CO-CHAIR CROOKS: This was intended to be a physician-level metric?

MS. CHRISTENSEN: Yes.

CO-CHAIR CROOKS: So, doesn't it still come out the same in a sense that, if you have 12 patients and one patient is 10 percent and one is 20 and one is 40 and one is 60, you can do a numeric average of those percentages, and it is going to tell you sort of the average number of months during the year that your patients were out of compliance or above that? No?

Okay, let's ask the developer to illuminate.

MR. JONES: I think your interpretation, Jeff, is it is the number of months. I cannot answer if you average, add

them up, an average, you come out with the total percentage of patients.

MS. LEBEAU: Can I? Just from a layman's perspective, it would seem to me that then you are getting at the chronically-high, the folks who are an ongoing problem all the time, as opposed to somebody who may stray over the line once. Now that may just be my intuitive sense of this, but that is what it seems like to me.

CO-CHAIR CROOKS: So, a target could be set, for example, that 10 percent is okay, which would be one month a year, but 50 percent is not good. Is that the way it is intended to be used as a safety measure?

DR. BERNS: I think a value, I mean the way I interpret this, again, and this gets to the issue that I raised the last time we met about having a measure related to, say, consecutive months above 12 rather than a month above 12, is exactly what you said. This sort allows forgiveness for your first

speeding ticket, but if you get three in a row, you know, you have to pay, you've got points. I think that is how this would translate into practice.

DR. NALLY: So, now I am a little confused. Let's say you have the 12 patients, and two are greater than 12 every month, and the rest of them are fine. As your stable of patients, you get an answer that seems fine.

I am concerned that you could potentially miss that signal for the patients we are trying to protect, which are those always over 12. So, the question is, is that measure protecting those patients?

DR. BERNS: Yes, I'm not sure that this does everything that you would want it to do. But, again, I think as you add it up, the way I would think of this, again, is if you have one patient who is above 12 for the whole year, it is 100 percent times one patient. A patient who is half the year would be 50 percent times one patient. And, then, you add

up all those, and you come up with a percentage.

There is going to be a lot missing there, as you say, because it will be an average over all of those patients. But it probably would give a signal. The higher this number, obviously, the more patients you have who are spending lots of time above 12.

Again, you could get at the same information by saying the percentage of patients who have hemoglobins above 12 for three consecutive months or six consecutive months, or what have you.

DR. KLIGER: Again, just real quickly, it is the reason that we advocated last time for calling a safety signal or a failure, people with consistently high calciums or consistently high hemoglobins. I think that that, Joe, would be a better way for us to do it, but we don't have that before us. And with the measure that is before us, I share Jeff's sentiment.

CO-CHAIR CROOKS: Okay. Are we ready to vote on the body of evidence?

DR. PACE: Okay. So, let's vote on quantity of evidence, high, moderate, low, insufficient.

Start the timer.

MS. RICHIE: And Lorien, quantity?

DR. DALRYMPLE: Moderate.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Oh, someone's out? Okay.

DR. PACE: Okay.

CO-CHAIR CROOKS: Go ahead.

So, we have 4 high and 17 moderate. Okay.

The next slide is the quality.

DR. PACE: All right.

CO-CHAIR CROOKS: All right.

MS. RICHIE: Lorien, quality?

DR. DALRYMPLE: Moderate.

MS. RICHIE: Kristine?

CO-CHAIR SCHONDER: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Waiting for one more, and you might revote. Someone is abstaining. Okay, we can go.

Go ahead.

All right, 1, high; 18, moderate; 1, low.

And finally, the consistency.

DR. PACE: Okay.

CO-CHAIR CROOKS: Go ahead to the next one, please.

Consistency of results across the body of evidence, high, moderate, low, insufficient.

Go ahead.

MS. RICHIE: Lorien, consistency? Lorien?

DR. DALRYMPLE: Moderate.

MS. RICHIE: Kristine?

CO-CHAIR SCHONDER: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, I think we are only going to get 20.

DR. PACE: Okay. Go ahead.

CO-CHAIR CROOKS: Go ahead and stop it.

Two, high; 16, moderate, and 2, low.

So, this would pass the quality of evidence.

DR. PACE: Right. Right. Okay.

So, basically, where are we at, Lauren? Impact was fine, opportunity for improvement, and now evidence. So, we can move on and talk about reliability and validity. Okay.

CO-CHAIR CROOKS: Ruben?

DR. VELEZ: On the reliability, again, we talked already about the numerator and the denominator, but we are talking again about calendar months on both ends, on the numerator and the denominator. And they talk

about consecutive months, which is the other thing.

I don't think, they don't have any risk adjustments on this measure. So, there is no risk adjustment.

And I really don't have any other comment at this time. So, I will open it for comments.

DR. PACE: Okay, and we will specifically talk about reliability and validity separately. So, right now, reliability. And risk adjustment, we can address whether that is an issue under validity.

DR. VELEZ: Okay.

DR. PACE: So, on the reliability testing, did you think -- it looks like the reviewers thought that it met the moderate category. So, were there any questions about reliability? Jeff?

DR. BERNS: So, again, since this is a measure that spans CKD 4, 5, not on

dialysis as well as on dialysis, I think we need to be very clear where this has been tested, because I suspect it has not been tested at all in the CKD not on dialysis or in enough venues to get any sense that this can be reliably collected and analyzed in different practice settings.

DR. KLIGER: So, I wanted to ask the developer where this was tested. Can you give us some information about the testing?

MS. CHRISTENSEN: Sure. I don't know if maybe you guys can bring up on this screen, the scientific acceptability section of the form. On mine, it is page 27 to 43, but I don't have your copy.

So, it kind of goes through the data sample. This was tested in four nephrology practice sites that represented a variety of types, locations, and sizes, to get a good cross-section of the environment.

Number of physicians per site ranged from 5 to 62 physicians in four

different regions of the United States. Patient volume ranged from 240 to 2800 ESRD patients seen per month. The sample size per physician organization ranged from 24 to 30 patients for patients on peritoneal dialysis or hemodialysis.

And we used the analytic method of both percent agreement and the kappa statistic, which adjusts for chance agreement. And this measure came out as highly reliable with a kappa of 0.9860 and 99.45 percent agreement.

DR. BERNS: These are all dialysis patients.

PARTICIPANT: I need CKD data. Are there any?

DR. BERNS: This is tested in all, in non-dialysis. This was not tested in non-dialysis patients?

MS. CHRISTENSEN: It was, but we presented the ESRD data because that was the primary population.

DR. BERNS: Well, except there's actually many more patients with CKD 4 and 5 --

MS. CHRISTENSEN: Yes.

DR. BERNS: -- in this country than there are on dialysis.

MS. CHRISTENSEN: Hang on one second.

DR. LATTS: And could somebody else maybe speak to the wisdom of having these as one measure with the two populations as opposed to two separate measures?

DR. PACE: I'll just mention from one standpoint, if it is the same target or the same numerator, there are some advantages to having one measure. You are always sure that they are going to be harmonized. It captures the intended population.

If you think there is a reason that it should be stratified, that can be discussed. I mean there are pros and cons to it, but those are some of the reasons.

DR. FENVES: Then, of course, my key question would be, how exactly is CKD 3 and 4, especially 3, defined?

DR. BERNS: This is just 4 and 5.

DR. FENVES: Okay. Well, I'm sorry. Even 4 defined, is it estimated; is it basically creatinine? Is it isometric clearance?

DR. DALRYMPLE: It is what is provided in the e-specification.

DR. PACE: Right, so that would be in the denominator specifications, which would be, let's see, for the denominator, let me find it.

DR. DALRYMPLE: And I don't know if this is a good time to talk about the e-specification, but I don't know if others also had concerns about some of the data elements listed and the patient population, including things like procedure codes for continuous veno-veno hemodiafiltration, which I think would mostly be acute kidney injury.

And then, under the different hemoglobins being pulled under laboratory tests, there were numerous hemoglobins, including hemoglobin F1, hemoglobin S, sulfhemoglobin. There were just numerous lab measures that weren't actually relevant to this measure, and I didn't know what the other Steering Committee members thought about the e-specification.

CO-CHAIR CROOKS: Go ahead and respond.

MS. CHRISTENSEN: Sorry. Yes, I am not responding to that, but back to the point, the question that I needed to look up for the CKD patients. The kappa was 0.9867, which, again, is highly reliable, and the reliability was 99.4 percent, agreement percentage.

CO-CHAIR CROOKS: So, let me clarify and ask you this: it doesn't say this specifically, but did you have raters go in and look at the same data? In other words, Rater 1 would look at Patient A data, and then

another rater would come and look it? That is what you are talking, inter-rater reliability? Okay. That's good. That is what consider to be reliability testing.

DR. DALRYMPLE: Can I clarify? Was the reliability testing done based on chart review at the facilities or how you are proposing to implement the measure using CPT codes and EHR data?

CO-CHAIR CROOKS: Well, that's not implementing. That's just the way --

DR. DALRYMPLE: That was a point of confusion for me.

DR. PACE: No, it is specified for CPT II codes.

DR. DALRYMPLE: Because the numerator details are going to be a CPT II code, correct? I guess when I read the reliability testing, I thought you were actually maybe going into the clinics and abstracting data straight from the charts as opposed to looking at CPT II codes. But I was

just hoping to get clarification on that.

MS. CHRISTENSEN: Excellent question.

So, the measure is available for us in a variety of data sources, including use CPT II codes, but also for EHR and for chart review. So, the inter-rater reliability, we did do with two human beings doing manual abstraction from either an EHR or a paper record.

We did compare to PQRI implementation, which is what I believe you are speaking of with the CPT II codes, where that was possible. It is on page -- sorry, there's a lot of data here.

Well, I can tell you it is about 60 percent for this measure. There was trouble with the CPT II coding. This was one of the first years that PQRI was implemented. And because it was a monthly measure, the way the facilities do their billing, which I am sure is not news to you guys, is on a monthly

basis. They have one charge for the month of care. And a lot of times, they were using the lab data from, for example, May if it was the June bill. So, that didn't match up with the way that the abstracters did that in calendar months. So, take what you want to out of that number.

DR. PACE: So, the testing was done on medical record chart abstraction, but you are not really intending to have the measure measured that way, right, going in and doing chart abstraction?

MS. CHRISTENSEN: I believe we submitted this for manual paper chart review, for EHR specifications, and for claims.

DR. PACE: And, Lorien, you had concerns about the EHR specifications? Is that what you were saying earlier?

DR. DALRYMPLE: Right. Just about trying to clarify some of the data elements. It was unclear to me some of the procedures codes that were being included, you know, like

continuous veno-veno hemodiafiltration and, also, all of the hemoglobins. Many are things like hemoglobin F, which is fetal hemoglobin, and it didn't seem relevant, and other hemoglobin variants.

And I know the exclusion criteria has sickle cell disease and other things, but it was unclear to me why so many different types of things that aren't relevant to the measure are being included in the e-specification.

DR. PACE: Okay. Do you have a response?

MS. CHRISTENSEN: We do not have a specifications person here. I am not sure if there is one available on the phone or not.

CO-CHAIR CROOKS: So, if I am understanding this right, if the data is coming from different sources, do we require that they check the validity of every single possible way of getting data? You know, chart abstraction is one. And they have shown that,

if you take the chart and you go over and put the data in the system, it seems to be pretty reliable. But there are other methods of collecting data, and you haven't checked the validity of every method separately?

MS. CHRISTENSEN: We did compare the PQRI data to the manual abstraction. But, like I said, the results are probably lower than they would be if that was tested again today.

CO-CHAIR CROOKS: Well, I would say this: that compared to a lot of the reliability testing we get on these forms, I give you great credit for having done it and actually reported it. And I think that, so --

DR. KLIGER: So, Pete, just a rejoinder to that, and I agree, is that that was only for a segment of the population we are being asked about reliability testing. It is only tested in ESRD patients, not in Stage 4 or Stage 5 CKD patients. We have no data for that.

MS. CHRISTENSEN: I actually read that. I'm sorry. I can read that again for you. It was maybe a .01 off of it. Let me find it again.

The CKD data was for the reliability percentage, the percentage agreement, 99.4, and the kappa was 0.9867.

DR. KLIGER: So, how did you test? What population did you test CKD, not on dialysis? How did you do that testing? How did you find those patients? How did you test this in non-dialysis patients?

CO-CHAIR CROOKS: Describe how you found the patients and how you did the study.

MS. CHRISTENSEN: Okay. So, there's denominator specifications for the measure, and the denominator specifications are found using the clinic systems to meet the specific codes or conditions that they are supposed to be on.

DR. KLIGER: Okay. So, in the physician practices, based on the CPT coding,

you found Stage 4 and Stage 5 CKD patients, and in that population you tested the validity --

MS. CHRISTENSEN: Yes.

DR. KLIGER: -- or the reliability? Thank you.

MS. CHRISTENSEN: Yes, and I apologize for not including both of them.

DR. DALRYMPLE: And this is Lorien again. I apologize for asking this again. I just want to make sure I understand correctly.

When the measure is actually implemented, there will be a component of chart review or it will rely only on CPT II codes?

CO-CHAIR CROOKS: Yes, go ahead.

MS. CHRISTENSEN: It would depend on how the program or the institution decided to implement the measure. So, they could implement it using claims. They could implement it using EHRs. Or, if they had no other way to do it and still wanted to do

quality improvement in this area, they could do manual chart review. Obviously, that is the least-efficient method.

DR. DALRYMPLE: Okay. Thank you.

CO-CHAIR CROOKS: Okay. So, let's vote on 2a, reliability, including the precise specifications and the reliability testing. High, moderate, low, or insufficient.

DR. PACE: Okay.

MS. RICHIE: And Lorien?

DR. DALRYMPLE: Low.

MS. RICHIE: Low?

And Kristine?

CO-CHAIR SCHONDER: High.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Steve is not back. Is anybody else missing? We are not getting the 21. Oh, got it there. Okay.

All right, we have 4, high; 9 voting moderate; 5, low, and 3, insufficient. I think moderate carries the day.

DR. PACE: Okay. So, validity.

CO-CHAIR CROOKS: Okay, on to validity.

Ruben, can you comment on the validity?

DR. VELEZ: I am still trying to find the hemoglobin stuff that were discussed here.

(Laughter.)

But, I'm sorry, go ahead. The question?

CO-CHAIR CROOKS: We're up to validity now.

DR. VELEZ: Validity?

CO-CHAIR CROOKS: How did you rate it?

DR. VELEZ: Let me go back. On the validity, at least on the report -- and I'm sorry, I'm looking at my computer here --

DR. PACE: It looks like the preliminary reviewers, three rated it moderate. We have three --

DR. VELEZ: Thank you, because that

is what I was looking at.

CO-CHAIR CROOKS: And the validity -- I'm sorry, Ruben.

DR. VELEZ: No, no. Go ahead.

DR. PACE: Maybe you want to mention what type of validity. Is it face validity or some other type of validity that they --

CO-CHAIR CROOKS: Is this with --

DR. PACE: 2b.

DR. VELEZ: The panel or this expert panel was used to do the access to face validity of the measure. And there were 21 members. You can see them on your last.

According to the expert panel, seven of them were either strong or very strong, 10 of them, on the testing results from internal validity.

Now that's it.

DR. PACE: Yes, that's fine. So, they did face validity, the measure score.

CO-CHAIR CROOKS: So, the question

asked to the panel was: "The scores obtained from the measure as specified will accurately differentiate quality across providers." And 17 out of 19 either voted 4 or 5, which is tend to agree or strongly agree, as their validity testing. Because they did reliability of their data elements, this is --

DR. PACE: Well, no, we don't necessarily combine them.

CO-CHAIR CROOKS: Right. But if they have done or if they claimed it was electronic, then we would like to see validity testing of the elements.

In a sense, though, what they did, is that also validity testing of the elements or just reliability?

DR. PACE: Primarily reliability of your abstracter. But, according to our criteria, face validity would meet the moderate rating. If you agreed with their assessment, if you had questions about it, then we would have to talk about it, yes.

DR. BERNS: A question: I don't see -- and this was raised before -- a definition of or how CKD 4 and 5 is determined. So, one issue that we should think about in terms of I think it's validity, although it might be reliability, is whether you used MDRD formula, which is what most commercial labs use, or whether you used CKD EPI, which labs may be using -- I'm sorry -- the question I am asking is, it is not specified in the denominator how CKD 4 and 5 stages, not on dialysis, are identified. And that patient population will be different depending upon the formula that is used, MDRD versus CKD EPI. It will also vary probably on the edges from month to month.

So, is a single estimated GFR that puts you in Stage 4 right on the borderline of 3 sufficient to flip you into this measure, and the next month you might flip out of the measure potentially?

So, those are just some

uncertainties I have about the denominator that may impact validity.

DR. PACE: Exactly. And that is a valid point to bring up. So, even though they did this, which is according to our criteria minimal on face validity, if there are concerns about the validity of the data to accurately capture the right patients, that is an issue for your discussion.

Before we vote on validity, you also need to address whether the exclusions in any way impact validity and, also, since this is an intermediate clinical outcome, whether there are any considerations that need to be reviewed regarding risk adjustment, or why not. So, all of that kind of factors into this ultimately, your vote on validity.

DR. VELEZ: And I may be somewhat confused, but going back and forth through this, I don't see any exclusions in this.

CO-CHAIR CROOKS: There are none.

DR. PACE: Right, but did they

identify some exceptions? Let me just look. It looks like none, right?

CO-CHAIR CROOKS: There's no risk adjustment, either.

DR. PACE: Okay. All right.

CO-CHAIR CROOKS: And no exceptions.

DR. PACE: Right. Okay.

MR. JONES: I believe I can answer the question about the categorization by stage. It was done by ICD-9 codes by the individual practice, but we don't know which formula that practice used to determine what stage that patient was in.

DR. PACE: And Lorien, you were mentioning in looking at the EHR specifications you had questions about, was it about the CKD codes or something else? I don't remember what you said.

DR. DALRYMPLE: When I was looking at the e-specifications, it does appear that all the CKD is based on ICD-9 coding. But to

identify dialysis patients, they also use a number of procedure codes. And some of the procedure codes, like continuous veno-veno hemodiafiltration, which is a modality we use in injury classically, not chronic dialysis, so there were some procedure codes that surprised me. And I didn't know how others felt about how the population was actually being identified, similar to the inclusion of all of these lab tests that didn't really seem relevant to the measure.

So, the procedure codes are on page 2 of the e-specification. And at least my understanding is this is how the initial population is identified, the IPP. And again, maybe these patients will fall out, but it is unclear to me why they are even being considered for inclusion in the IPP.

DR. VELEZ: Again, I don't find that data here. So, I am not sure we are talking about the same measure, but --

DR. DALRYMPLE: It is the coding

spreadsheet for PCPI e-specification AKid-7, patients on erythropoiesis stimulating agent, hemoglobin level greater than 12.

It was the appendix, correct, Karen? These are all appendix materials?

DR. PACE: Yes. In the folder with the measure submission form, there was an appendix, a PDF file. Lauren has got it up now.

And what page do you want us to take a look at, Lorien?

DR. DALRYMPLE: The initial pages are just kind of their outline of the flowsheets. But if you get to the actual, it looks like an Excel spreadsheet, where they start listing how the IPP is selected, what the numerator and denominator include, that is page 1 of that Excel spreadsheet, coding spreadsheet for PCPI e-specification.

CO-CHAIR CROOKS: We're looking at that now.

DR. DALRYMPLE: And that's where

you can see that it looks like all codes are being used for CKD Stage 4 and 5 identification. But as you scroll to page 2, at least my understanding is there are some procedure codes being used to identify dialysis patients. But some of these procedure codes include things like continuous veno-veno hemodiafiltration and extracorporeal albumin hemodialysis, and things that just are I don't think relevant to chronic outpatient hemodialysis.

And then, these are the same pages that include all of those different hemoglobin measures I mentioned that would appear to show up in the denominator. And that is further down on page 5.

So, my concern is, why are some of these being included in the e-specification?

CO-CHAIR CROOKS: You are arguing it may not be as valid as we think because there are patients who are getting acute dialysis procedures or other types of

procedures --

DR. DALRYMPLE: That we are not really interested in.

CO-CHAIR CROOKS: -- that we're not interested in. And in fact, I suppose could -- well, okay, I'll stop there.

DR. DALRYMPLE: They probably all have low hemoglobin. So, again, you could argue the relevance. But, again, this was one of my concerns under reliability when we were talking about specification of the data elements. Especially as you get to all the hemoglobins listed on page 5, you know, to include things like hemoglobin F1 and hemoglobin G and sulfhemoglobin, I mean these just are not laboratory measures that are relevant to the proposed measure.

CO-CHAIR CROOKS: Michael?

DR. FISCHER: Yes, I was going to say I think this is a big concern for the validity of defining the denominator. We have actually tried to look at this with VA data.

We did a study with ESRD, and we could not really come up with any algorithm of CPT codes for dialysis that would sort out chronic dialysis from acute dialysis.

Then, there are these continuous codes, which obviously aren't germane to a chronic population. But using CPT codes to identify chronic dialysis, at least our experience has been it was very problematic. And in the end, we had to use USRDS data to definitely define someone as having ESRD.

But I didn't review this directly. It wasn't assigned to me. I don't know if the people who created this measure, the measure developers have a response or if they had a particular reason why this was their approach.

CO-CHAIR CROOKS: Well, let's ask.

MR. JONES: I may be shooting in the dark here. I mean I am not a specification expert here.

But I think if you look at this, if

you go to the page where it says, "PCPI e-specification", it lists the CPT codes that were used. If you go to the table, all of those CPT codes are things that we would equate with ESRD. On that whole table are other things that were previously mentioned, the CBDH code, but those were not listed as a CPT code. They are in that table, but they weren't the ones that were used when the patient was categorized.

DR. PACE: Then, there is a mismatch between -- so, I guess part of the issue is, then --

DR. DALRYMPLE: Well, I think, are we talking about the CPT II codes or the procedure codes, the CPT II code to identify the numerator versus these procedural codes being used to identify processes? I may not understand the distinction.

I was thinking the CPT II code is going to potentially be used by some practices to identify the numerator. But, then, these

are going to be used, these procedure codes, to try to -- I think these are procedure codes. Others, I think, have more expertise in these, as to how you identify continuous veno-veno hemodiafiltration.

I thought this was going to be used to identify the IPP. But I am not an expert in this, either. So, I would definitely appreciate others' thoughts on how they interpret these tables and, then, what is included in the text.

DR. PACE: So, you want to clarify what the intention of this table was? We think it is your specifications for an electronic health record measure.

DR. DALRYMPLE: Uh-hum.

DR. PACE: But it doesn't seem to match what your English language denominator is.

MS. AST: No. Right, it is meant to be the EHR specifications. And like we said, we don't have a specification staff

person here. But I just got word that she is on the phone, but she cannot be heard.

So, she told me to say we appreciate the feedback and would be happy and willing to revise. I think it is the issue that the gentleman over here talked about. It is difficult to differentiate between the ESRD categories, but we are more than willing to revise, if there are some incorrect codes in there.

DR. PACE: So, I think one of the things we can do to move forward is we could perhaps divorce the EHR measure specifications from this measure at this point. And then, if they can bring in EHR specifications and show a crosswalk to the actual measure, then we can consider that being part of the endorsement. Would that work for people, if at this point we focus on the measure with the CPT II codes or the medical record abstraction process, divorce the EHR specifications at this point, move forward, and then we can talk with the

developer about whether it is possible to get EHR specifications in our timeframe? Okay?

CO-CHAIR CROOKS: Yes.

DR. PACE: Any objections to that approach?

CO-CHAIR CROOKS: And the Committee understands that there is no perfect method of identifying dialysis patients. We hope it will be as good as it can be. But perfection isn't the goal.

DR. PACE: Right.

CO-CHAIR CROOKS: It isn't the requirement.

DR. PACE: Right.

DR. BERNS: Can I ask one other question of the developer? That is the accuracy of the CPT -- I don't know whether it is the CPT or ICD-9 designation for CKD 4 and 5, whether that was done accurately. In other words, did you go back and confirm that, if the chart said CKD 4, that it, in fact, was CKD 4?

MS. CHRISTENSEN: Yes.

CO-CHAIR CROOKS: All right. So, we could --

DR. WELCH: Excuse me. I'm sorry.

CO-CHAIR CROOKS: Go ahead, Janet.

DR. WELCH: Can you just clarify? I think you said a few minutes ago that, for the purposes of the evaluation of validity in this piece, that face validity was considered evidence of moderate?

DR. PACE: It will meet our moderate, right.

DR. WELCH: Okay. That's what I -- because that is different than what I teach. So, I just wanted to --

DR. PACE: Yes, yes.

But one last thing before we vote on validity is any discussion about the fact that there is no risk adjustment. And the question is whether there is any analysis. Do you expect this intermediate clinical outcome to vary based on patient characteristics, you

know, hemoglobin level? And is there any reason to think that it would be significant enough to warrant some analysis? They didn't provide any analysis. So, it is just a question to the Steering Committee whether there is any question or issue that you want to bring up before we move forward.

(No response.)

Okay. Well, then, let's go ahead and vote on validity. High, moderate, low, insufficient.

MS. RICHIE: And Lorien, validity?

DR. DALRYMPLE: Moderate.

MS. RICHIE: Moderate?

Kristine?

CO-CHAIR SCHONDER: Moderate.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: That should be at 21, right?

DR. PACE: Okay.

CO-CHAIR CROOKS: Okay. Moderate,

15; low, 4; insufficient, 2.

DR. PACE: Okay. So, why don't we just move on and finish this one out, rather than breaking?

So, it would pass scientific acceptability. So, let's go on to usability. Oh, I'm sorry, disparities. We need to do a rating of high, moderate, low, or insufficient.

DR. LATTS: And again, just for the future, it would be nice to have an "NA" when you vote on this in the "thingamajiger".

DR. PACE: Okay. Good. Yes. Thank you.

CO-CHAIR CROOKS: Okay, usability.

DR. PACE: Okay, usability. Ruben?

DR. VELEZ: No, I don't have any issues. I mean they report here the usability in the sense of it has been used for the PQRI and PQRS in the past.

Again, I think we have to be careful because some of this, as was stated,

was dialysis more than CKD. According to their note, it has been also used in the 2009 and 2010 CMS PQRI programs.

I think, if I am reading this correctly -- oh, there you are. Two moderates and -- yes.

DR. PACE: And just to clarify, it is being used in those programs, but currently there is no performance data on physicians publicly available. So, it is being reported, but there is no access to performance data.

Okay. All right. Yes?

MS. CHRISTENSEN: We did present the 2008 data that is noted in here that it is confidential, but we were able to provide that to you. So, that is where the gap-in-care data came from.

DR. PACE: Okay. Ready? Any discussion about usability?

(No response.)

This is both for public reporting and quality improvement, and high, moderate,

low, insufficient.

Tenee?

MS. RICHIE: Lorien, usability?

DR. DALRYMPLE: Moderate.

MS. RICHIE: And Kristine?

CO-CHAIR SCHONDER: Moderate.

CO-CHAIR CROOKS: This is for public reporting.

DR. PACE: Usability for both public reporting and quality improvement.

CO-CHAIR CROOKS: For both? Okay.

(Whereupon, a vote was taken.)

DR. PACE: Okay, how many should we have this time? Oh, I think there's two people out.

CO-CHAIR CROOKS: Okay.

DR. PACE: Okay. All right.

CO-CHAIR CROOKS: So, for usability, 2, high; 13, moderate; 3, low; 2, insufficient.

So, it passes.

DR. PACE: Okay. So, we will go on

to feasibility.

And Ruben, anything to report?

DR. VELEZ: Not much. You can see that on the multiple questions most of the information is generated from provision of care, mostly electronic health records, and they are not aware of any unintended consequence at this point.

And the Committee voted anywhere between high to moderate on most of the questions.

DR. PACE: Okay. Any discussion about feasibility? Jeff?

DR. BERNS: I hate to beat a dead horse here, but just to clarify, feasibility can be looked at as, once you have identified the patient and you have their lab data, can you create the numerator and denominator? That seems to be what they are addressing here as opposed to feasibility is sort of taking a step back and making sure you have identified the right patients in a practice and

identified the physician responsible. So, there's several steps that need to occur prior to identifying a CKD patient or knowing that this is a CKD patient and having a hemoglobin level matched up to that.

I would like to just have clarity that that was what was addressed and that it is feasible across a variety of different practice settings, electronic health records. There's EPIC, there's Sunrise, there's paper. And all of those would need to be perused for this data. I am not sure we have information on the feasibility of that.

DR. PACE: Right. So, it is a good question. We would really expect to address most of what you mentioned under validity. Can you capture the data accurately? Can you have a valid measurement?

Under feasibility, the focus is more on burden of measurement, whether there are systems to capture the data. And the way we tend to think of this is that, because it

is exactly what people would bring up before, is that, well, it is usable if it is reliable and valid or it is feasible if it is reliable and valid. And we try to make distinctions there.

So, I understand that those things kind of carry over into the following criteria, but it really is more about that it can be captured and the burden of collection.

DR. BERNS: I guess my question, because I am seeing any evidence of feasibility.

DR. PACE: So, Jeff, do you have a specific question? Or are you just making a note that we really don't have information about feasibility?

DR. BERNS: Well, I guess I am asking if there is any data and then making the comment that I don't see any and not every practice in the United States has an electronic health record.

DR. PACE: Right. And basically,

the way it was tested, and so far the way it has been implemented, have not involved electronic health records. It has involved either CPT II codes off of claims or medical record abstractions. So, CPT II codes off of claims would also be an electronic source, but it is not an electronic health record exactly.

So, does PCPI have anything additional to say about feasibility?

MS. CHRISTENSEN: Yes. I am not sure if this is answering your question. So, please ask followup ones if it doesn't.

The way we did our testing was that we had the practice generate a list of patients they believed were eligible for the measures, and then the manual reviewers confirmed the denominator, the numerator, and exceptions, if there were exceptions, for the measure, independently of whether or not the other reviewer felt that way or the original list. They independently confirmed that.

Then, secondly, we did provide data

with the number of physicians that were using this in the PQRI program for 2009, which is, of course, PQRS now. The numbers have been going up year by year for implementation.

CO-CHAIR CROOKS: Does that answer your concerns, Jeffrey?

DR. PACE: Okay. So, let's go ahead and vote on feasibility, high, moderate, low, insufficient.

MS. RICHIE: And Lorien, feasibility? Lorien?

DR. DALRYMPLE: Low.

MS. RICHIE: Kristine?

CO-CHAIR SCHONDER: High.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, that should be it.

DR. PACE: Yes.

CO-CHAIR CROOKS: One, high; 10, moderate; 8, low; 2, insufficient.

DR. PACE: Okay. All right. So, let's move on to the last question for this

measure, which is now, overall, do you feel that the measure meets the criteria to be suitable for endorsement?

And again, this would be preliminary, and we would have to look at if there are any harmonization or competing measures issues. But if there weren't, then a yes vote would mean it would be recommended.

CO-CHAIR CROOKS: And, also, that they will provide more data on the validity questions that were outstanding.

DR. PACE: Well, what we talked about -- right now, we would be voting on the measure excluding EHR specifications until they would bring that back.

CO-CHAIR CROOKS: Right.

DR. PACE: Okay. Any questions, issues?

(No response.)

Okay.

CO-CHAIR CROOKS: The choices are yes, no, or abstain.

DR. PACE: Okay.

MS. RICHIE: And Lorien, endorsement?

DR. PACE: Well, suitable.

MS. RICHIE: Suitable for endorsement, yes, no, abstain?

DR. DALRYMPLE: So, this is suitable for endorsement, divorcing the EHR specification?

DR. PACE: Correct.

DR. DALRYMPLE: No.

MS. RICHIE: Kristine, suitable for endorsement?

CO-CHAIR SCHONDER: Yes.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, we have 21 responses.

So, we have 15 voting yes and 6 voting no.

DR. PACE: Okay.

CO-CHAIR CROOKS: Which is interesting because it passed all four

categories, you know. So, it should have been -- but here we have it.

(Laughter.)

DR. DALRYMPLE: Actually, can I ask, when we vote the yes/no, do we do it based on majority vote for each or on our personal vote for each of those, importance, reliability, validity, et cetera?

DR. PACE: What's your question?

CO-CHAIR CROOKS: What's your question?

DR. DALRYMPLE: When we vote whether to endorse the measure, you know, I'll just speak for myself. I voted based on how I had rated the criteria. So, for example, because reliability and validity were low, that would be a non-pass for me personally.

CO-CHAIR CROOKS: Yes, that's valid. That's certainly valid. If that is consistent with your assessment of the measure, then that is certainly fine.

DR. PACE: Yes.

DR. DALRYMPLE: Okay. So, we don't vote how the majority is voting for each of those.

CO-CHAIR CROOKS: No.

DR. DALRYMPLE: You know, the majority passed on validity and reliability. But if we didn't personally pass on those, it is okay to say, well, on my rating they didn't meet the criteria?

CO-CHAIR CROOKS: Yes.

DR. PACE: Yes.

CO-CHAIR CROOKS: You're right. That's fine.

DR. PACE: Yes, absolutely.

CO-CHAIR CROOKS: And my comment was that I guess, typically, we had seen that, if it passed the other ones, it would go through more easily. But you certainly should vote consistent with your assessment of the measure.

Okay. I think we have reached lunchtime, haven't we? I think us West

Coasters need some protein to proceed.

DR. PACE: But we need to do --

CO-CHAIR CROOKS: Oh, that's right, we need to have a public comment period before we break.

DR. PACE: One more second before we rush to lunch.

CO-CHAIR CROOKS: So, public and metric submitters.

DR. PACE: Right. So, let's go to -- is there anyone on the phone that wants to make --

CO-CHAIR CROOKS: First, on the phone. Any public, non-metric submitters who would like to make comments? Or anybody else?

Yes?

DR. ASHFAQ: This is Dr. Ashfaq. I'm in the Clinical Development in Amgen.

I would like to comment on the sub-10 measure, even though you have voted on that measure, but I would like to comment anyway.

We are very concerned about not

monitoring this sub-10 measure. I would just like to point out the clinical evidence part we discussed here. We have clinical evidence from our registrational trials that sub-10 decreases the transfusion. That is how the drug came in the market.

We also have clinical trial evidence from normal hematocrit study showing the difference in transfusions in patients who were targeted at lower hemoglobin levels versus high.

We also have very robust government-funded data from USRDS which is tracking these measures for a long, long time. And I am just going to give you an example.

In 1991, when sub-10 was 60 percent, the transfusion rates per quarter were 14.4 percent. When the hemoglobin sub-10 dropped 5 percent in 2001, the transfusion rates decreased to 8 percent.

We also have data to suggest that there are other outcomes which may be adverse

associated with hemoglobin less than 10. For example, patient-reported outcomes, we have clinical registrational trial data which suggests that exercise tolerance and ventricular function is impacted.

But the normal hematocrit data as well as the registrational data is actually presented in the modified FDA label which was recently published.

We also have an abundance of data, associative data, also including the DOPPS data, suggesting that hospitalization increases with hemoglobins less than 10.

I think what we are doing is that the pendulum has swung from the safety on the higher end to now on the lower end. And I think we are going to be reactive if we are not going to monitor these sub-10s. We want to be proactive. We would like to continue to monitor hemoglobins less than 10.

In the meantime, we should come up with more robust measures which are related to

outcomes, including transfusions. And while we are doing this, we should continue to monitor hemoglobins. So that, when we have those robust measures, we can couple with the hemoglobins.

Thank you.

CO-CHAIR CROOKS: Thank you.

Any other comments? Yes?

You need to get to a microphone. So, you could sit down there.

MS. SCHUBERT: Very quickly, my name is Katy Schubert. I am the American Society of Pediatric Nephrology's Washington representative.

I just wanted to voice ASPN's support for Measure 1667, which is coming up after lunch.

While there is limited research on pediatric ESRD patients receiving dialysis having hemoglobin levels less than 10, studies that have been done have shown a 60 to 70 percent decreased risk for mortality among

adolescent patients with hemoglobin greater than 11.

Additionally, anemia in children on dialysis and with chronic kidney disease has been found to impact negatively several aspects of health-related quality of life. This topic has been given a priority nomination for the 2012 Best Pharmaceuticals for Children Act with the National Institute for Child Health and Human Development.

And we believe that the target is appropriate for the pediatric and adolescent patient population. More generally, ASPN does see the need for more pediatric ESRD and chronic kidney disease quality measures at both the physician and the facility level, and we support harmonization on analogous measures when that is appropriate.

We will continue to work with the AMA PCPI and RPA in the area of physician-level measurement in children as well as collaborate with CMS on facility-level

pediatric measures, and then, moving forward with the implementation of the QIP for ESRD facilities, which will include pediatric measures in the future.

We believe that NQF endorsement of Measure 1667 may lead to this measure's inclusion in the PQRS, which will further the goal of giving the best quality of care for this pediatric population.

Thank you.

CO-CHAIR CROOKS: Thank you.

Any other comments? In the back?

MS. McGONIGAL: Hi. I will be brief, so you guys can get to your lunch.

I'm Lisa McGonigal from Kidney Care Partners, which is a national coalition of patient advocates, healthcare professionals, care providers and suppliers, working together to improve care for patients with chronic kidney disease.

We are also the convener of the Kidney Care Quality Alliance, which developed

some of the measures that you will be discussing tomorrow. And we do appreciate this opportunity to comment on the measures that you are considering here today and tomorrow.

For the anemia management measures, we wanted to voice our support for Measure 1666. We would like to support this for both reporting and payment purposes. KCP's position on the other three anemia management measures is to support them for public reporting only, not for payment.

Also, on Measure 0252, given the performance gap between the HD and PD patients that was mentioned previously, if this measure does eventually become endorsed, when implemented, we would recommend that it be reported separately by modality.

I would also like to take one minute to comment on this afternoon's session prospectively. For the cardiovascular measures, we support the following measures

for public reporting only: 0627, 1662, and 1633. We support 1668, lipid profile testing, for both public reporting and payment, and we continue our prior opposition to 0626, CKD lipid profile monitoring, because it is not harmonized with the corresponding PQRI measure and it is not strictly consistent with the KDOQI dyslipidemia guidelines.

Finally, for the dialysis adequacy, KCP previously supported the process Measures 0247, 0248, 0253, and 0254, but we are now recommending that these be retired, in the interest of endorsing a parsimonious set and given the availability of corresponding adequacy outcome measures.

We support the following dialysis adequacy outcome measures for public reporting and payment: 0318, 0321, 0323.

And finally, we previously supported the outcome Measures 0249 and 0250, which are the minimum delivered HD dose at greater than six months and greater than 90

days. However, we note that the knowledge base has evolved since the measures' initial endorsement in 2008, with recent data suggesting that longer treatment for incident patients might reduce 90-day mortality rates, rendering the residual renal function exclusion unnecessary.

We, therefore, recommend that a single minimal delivered HD dose measure be used, specifically 0249, but that the measure should commence on day one of dialysis rather than at six months, and there should not be an exclusion for residual renal function. And we would support this amended measure for both public reporting and payment.

Thank you.

CO-CHAIR CROOKS: Okay. Any other comments?

Thank you.

DR. ASHFAQ: Sorry.

CO-CHAIR CROOKS: One more?

DR. ASHFAQ: Just one thing. We

have submitted a lot of data during the MEDCAC to support the sub-10. But if the Committee is interested, we will be more than glad to submit it to you, too.

CO-CHAIR CROOKS: All right.

When we come back, we are going to -- or do we need to know this right now?

DR. PACE: No, if you would just let Lauren know, yes.

CO-CHAIR CROOKS: Okay. For those who are interested in dinner tonight, it is still a open possibility. So, let Lauren know. Okay.

DR. PACE: Okay. So, let's get a well-deserved break for lunch. Obviously, we are behind schedule, and we will see if we can make some time up this afternoon and possibly go a little bit longer than we had planned today.

But let's try to get your lunch and reconvene in about 20 minutes, so that we can --

CO-CHAIR CROOKS: So, we can do a working lunch.

DR. PACE: Right.

CO-CHAIR CROOKS: We can eat while we are --

DR. PACE: Right.

CO-CHAIR CROOKS: Okay. So, 20 minutes, then we will try to resume again.

DR. PACE: Right.

CO-CHAIR CROOKS: Which will be quarter to 2:00.

DR. PACE: Yes. Right.

(Whereupon, the above-entitled matter went off the record at 1:26 p.m. and resumed at 1:55 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:55 p.m.

CO-CHAIR CROOKS: Okay. The next measure is 1667, pediatric ESRD patients receiving dialysis, hemoglobin level less than 10.

Presenting, Rick? Dr. Kaskel?

DR. KASKEL: Okay. So, this measure is evaluating the percentage of calendar months within a 12-month period during which patients aged 17 years and younger with a diagnosis of ESRD receiving hemodialysis or peritoneal dialysis have a hemoglobin level less than 10 grams per deciliter.

There's a number of exclusions here, the same as were seen in the 1660 measure except, in addition, hypersplenism was added as well as post-operative bleeding, active bloodstream or peritoneal infection.

It is an outcome measure, and the data source is the same as 1660. It is not

paired, nor is it a composite.

Just as a review, we heard before that CMS has withdrawn the pediatric facility-level measure of a hemoglobin less than 10 recommended in the recent project, due to the FDA announcements.

The evidence provided for reliability and validity of this measure appear to be the same as that which was presented for the adult Measure 1660. But I am going to try to review some unique criteria and evidence-based medicine for pediatrics that show this is different than what we just heard in 1660.

Can I proceed to give a little review?

DR. PACE: Why don't we start with, Tenee, 1a? So, we will go through like we did before and start with impact, 1a, high impact.

And Lauren has got the preliminary results up. I don't know, Rick, if you want to say anything about that. Basically, go

ahead.

DR. KASKEL: The group appeared to be in uniform agreement with three high and two moderate.

DR. PACE: So, do any of the other reviewers or Committee members want to make any comments about impact or are you ready to vote on that?

(No response.)

Okay. Let's go ahead. Impact for this measure, high, moderate, low, insufficient.

Go ahead.

MS. RICHIE: And Lorien, impact?

DR. DALRYMPLE: Moderate.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: I think we are missing a couple. So, I think 18 will be it.

Okay. All rated high or moderate, 8, high; 10, moderate.

DR. PACE: Okay. So, now let's go

on to opportunity for improvement or performance gap.

DR. KASKEL: For opportunity for improvement/performance gap, we see we have almost uniformly four moderates and one insufficient.

DR. KLIGER: Rick, could you give us the data for that, for the performance gap?

DR. KASKEL: Surely. We have evidence from two major sources. One is a review from the Children, the North American Pediatric Renal Transplant Cooperative Study Base, from 1992 to 2001, with hemoglobins less than 9.9 grams per deciliter compared to those with hemoglobin values greater than 9.9 grams per deciliter. And this showed an elevated risk for mortality and greater risk for hospitalization in the groups that had the lower hemoglobin levels.

In addition, a more recent report from NaProTech's looked at over 2,079 patients ages two years and older with CKD Stages 2 to

5 found the prevalence of anemia in this group as defined as a hematocrit less than 33 percent, had increased significantly between Stages 2 to Stage 5. So, anemic children were also 55 percent more likely to be hospitalized when compared to non-anemic children with CKD. That is a more recent report.

DR. KASKEL: So, right. I appreciate those correlations, but I wonder about the performance gap. Do we know how many patients in fact don't achieve the level that is stated in this measure?

DR. PACE: So, Lauren, do you want to pull up 1b, which they presented?

DR. KASKEL: That's adult children.

DR. KLIGER: Data for children for the performance gap, is that what you are saying?

DR. KASKEL: That's the older data. Nothing new exists.

DR. KLIGER: No, no, I understand. So, for performance gap, we don't have any

data, correct?

DR. KASKEL: Correct.

DR. KLIGER: Okay. Thanks.

DR. PACE: Okay, measure developer?

MS. CHRISTENSEN: It's in 1b2 of your forms. The gap in care shown by the PQRS data in 2008, 36.51 percent of patients reported on did not meet the measure.

DR. KLIGER: That is pediatric patients?

DR. DALRYMPLE: It looks identical to the adult data.

DR. PACE: Right. That's the question. Is this --

MS. CHRISTENSEN: That's a good question. I don't know the answer to that.

DR. KLIGER: It's not pediatric patients?

DR. PACE: Yes, this is the adult data.

DR. KLIGER: Yes. So, let's just be clear. For this pediatric measure, we

don't have any performance gap data.

DR. PACE: Right. Okay. And I guess the question is, is anyone on the Committee aware of existing data on performance gap for pediatric patients?

DR. KASKEL: We don't have data on the performance gap. We have some data, recent publications showing some of the adverse outcomes of anemia in this population.

DR. PACE: Right. Okay. We will get to that with evidence. So, okay.

So, let's, I guess, go ahead and vote on this. Performance gap would be -- and before we -- well, let's go ahead and vote on this.

DR. LATTS: Can I just ask, do we know why we don't have performance gap data? Is it not being collected? Or do we just not have it?

DR. PACE: Right, because this doesn't have to be from the measure as specified. It can be from the literature,

from prior studies about what percentage of kids have hemoglobins below 10. There's nothing like that in the literature, anyone?

DR. KASKEL: We don't have any review of any recent update of that, other than what I have presented before. It's lacking.

DR. SOMERS: Well, I mean, there is data in the literature to suggest that there is a proportion of children who have hemoglobins less than 10 that I think exceed the adult number, from my recollection of that. I can check to see if I actually have something here.

DR. BERNS: Can I ask, maybe it is a silly question, reflects my ignorance about this. But, as we are thinking about the importance, the impact of this measure and other pediatric measures, how many kids are on hemodialysis in the United States?

DR. KASKEL: We're upwards of about 2500 to 3,000 total, maybe a little more than

that.

DR. SOMERS: Around 2500, yes. Yes.

DR. KASKEL: We do have data. We do have recent data that came out of a Chronic Kidney Disease in Children Study showing that over 40 percent of children with Stage 2 to 4 CKD in North America are anemic. But that didn't define the level of hemoglobin, the percentage of that have hemoglobins below 10.

DR. LATTS: And I do recall from last time that Barbara had said, I think, that they mostly get dialyzed in specialty centers, that they are very concentrated.

DR. SOMERS: Yes, that is correct, especially children who are getting hemodialysis, yes, and younger children.

DR. PACE: Okay. Any other discussion about opportunity for improvement or performance gap?

(No response.)

Any other information anyone wants

to share?

(No response.)

Okay. So, let's go ahead and vote. High, moderate, low, insufficient.

MS. RICHIE: And Lorien, performance gap? I'm sorry, what was that?

DR. DALRYMPLE: Insufficient.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Is that it? Twenty? Yes.

Okay. Nine voted moderate; 11 said insufficient.

DR. PACE: Okay. Let's go on, then, to evidence. This is not a health outcome. It is an intermediate clinical outcome. And we will be talking about the quantity, quality, and consistency of the evidence.

I think we need to talk about this in light of our prior discussion about the evidence, and I think, Rick, you think that

there is a difference for pediatric.

And one other context is this Committee did recommend a similar measure for facilities last time, but CMS withdrew that measure.

Okay, Rick?

DR. KASKEL: For quality of body of evidence, determination of hemoglobin targets in pediatric patients really resists definitive recommendations. The quality of life is important, obviously, to the development of the child and their family. This leads urgency to the consideration of higher hemoglobin thresholds. Age-specific variation also in normal hemoglobin levels introduces further uncertainty. And given the metabolic growth and needs and psychosocial differences between children and adults, we need to rely particularly uniquely on pediatric data.

There has been a couple of studies that should be updated. A recent study in

2006, a randomized controlled trial by Amaral, did show evidence for the benefit of treatment of anemia with ESA versus placebo, such that hemoglobin levels greater than 10 grams percent in children were associated with a partial correction of an elevated cardiac index by six months of therapy and a reduction in left ventricular mass by 12 months.

A second study by Garrison in 2004 looked at 105 pediatric hemodialysis patients and found that those who had hemoglobin levels less than 10 grams percent were associated with poor quality-of-life evaluations and poor performance, both physically and in school.

So, we have two recent studies that would suggest that treating the anemia is very important.

And finally, just an update to what was mentioned before, we have so little data in pediatric CKD and ESRD that a nomination application to the Best Pharmaceutical for Children Act was put in place this winter and

was prioritized by the BPCA and the NICHD for the 2012 priorities for studies in pediatrics. So, it received one of the several priority scores to have further research done on this important issue, including targets of ESA treatment.

CO-CHAIR CROOKS: Is it fair to say, though, the body -- and this kind of comes up again and again -- the body of evidence may strongly support the importance of treating and that patients can improve if addressed in some outcomes? But, just as in performance gap, there is nothing really tied to the frequency of measuring the hemoglobin?

DR. PACE: This is about less than 10.

CO-CHAIR CROOKS: Oh, this is less than 10? I'm sorry. Okay. So, let me withdraw my question and take another comment.

Alan?

DR. KLIGER: Mike and Rick, I am just interested in your opinion about the

quality of those two studies. We haven't had a chance to review them, but they both specifically address the hemoglobin target that we are talking about here with specific outcomes that are more pertinent to children than to adults. So, can you give us your assessment of the quality of those two studies?

DR. SOMERS: Sure. Sure. So, the Amaral study that Rick alluded to looked at almost 700 kids, and it used Clinical Performance Measure Project data linked with USRDS hospitalization and mortality records. It really did show that, in terms of mortality, there was like a 70 percent difference if your hemoglobin was less than 10 versus greater than 10. There was also a significant difference in rates of hospitalization as well.

So, I mean, I think that that, for the pediatric world where we have small numbers and we are stuck with very limited

data, that, for us, is very strong data to support deleterious consequences of hemoglobin less than 10.

In addition to that, as Rick alluded to, there are smaller studies showing improvement in measures of cardiac health as your hemoglobin goes greater than 10 as well.

Then, there is data from a cohort of about 150 or 160 kids looking at a validated measure of quality of life and looking at how anemia negatively impacts health-related quality of life, and especially measures that in children are important in terms of physical development, cognitive development, school attendance, school performance, as well as social interactions with family and friends as well, again, showing that, as you become more anemic, you have a much poor quality of life.

DR. BERNS: What is the pediatric nephrology world take on this? We sort of were led astray in the adult patient world by

retrospective observational studies. So, what is the thought among pediatric nephrologists about what we found to be this discordance between prospective and retrospective studies? And do you think that applies to kids? I am just curious because a lot of this is going to be guided by this one study probably.

DR. KASKEL: We have a very successful prospective evaluation going on, a longitudinal cohort study. It is not a treatment study. But it is a longitudinal cohort study, ongoing assessment, now into its third round of funding. It is similar to the CRIC Study, and we are looking at children not on dialysis but Stage 2 to 4. And then, as they transition to dialysis, they are in another study.

But that has yielded very new and provocative information about the factors that are unique to pediatrics in CKD. We have found that 40-odd percent of them are anemic. Another 40 percent have hypertension that is

masked hypertension in that population. They have normal blood pressures in the clinic, and on 24-hour inventory they were abnormal. And those 40 percent that had masked hypertension, a significant number had LVH. So, we're learning things.

And as far as the anemia is concerned in that study, it is begging a trial to determine why there is such a high percentage of anemia in children. Again, you have the confounding factors of growth. Age-related differences in hemoglobin have been shown in the normal population. When you have impairment of growth in CKD, it is a whole host of other factors, nutritional, hormonal, that are working. And, then, the micro-inflammation that many of these children have demonstrated, again, in early stages.

So, I think we have a lot of room to move ahead with the appropriate anemia management. I don't believe that we are facing the same issues that were seen in the

adult studies looking at excess hemoglobin targets. It is inadequate hemoglobin targets for the child.

CO-CHAIR CROOKS: A lot of nodding going on. So, I think we are ready to --

DR. PACE: Right. Would you just clarify, though, I know in the submission it mentioned that the recommendation is still considered opinion-based, expert opinion-based.

DR. KASKEL: Yes. We don't have the trial to define it.

DR. PACE: All right.

DR. KLIGER: These are retrospective? You are talking about these data are what you think of as well-done, observational, retrospective studies? So, that is why you are calling it opinion-based?

DR. KASKEL: That's right. And Michael said that one of the studies was very well-planned.

CO-CHAIR CROOKS: Are you sure you

would call that opinion-based then? Because you are saying there is a lot of evidence and well-done studies; they are not all clinical trials, but that is not the only kind of --

DR. KASKEL: It's evidence-based, yes.

CO-CHAIR CROOKS: It's evidence-based.

DR. PACE: I was just referring to the submission form talked about expert-opinion-based. So, that is a question for you all.

Okay. So, any more discussion about the evidence that does or does not exist for the less-than-10 target?

(No response.)

Okay. So, we will first rate quantity, high, moderate, low, insufficient.

MS. RICHIE: And Lorien, quantity? Lorien?

DR. DALRYMPLE: It's moderate.

Can you hear me?

MS. RICHIE: Yes, I can hear you now. Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, we're up to 19. That sounds about right. We have 20. Okay. Great. That's got to be it.

All right. Seventeen, moderate; 1, high; 3, low; 1, insufficient.

Let me do it again. One, high; 17, moderate; 1, low; 1, insufficient.

DR. PACE: Okay.

CO-CHAIR CROOKS: Okay. So, we can go on to the quality.

DR. PACE: Uh-hum. All right. Quality rated on high, moderate, low, insufficient.

CO-CHAIR CROOKS: Go ahead and start that.

MS. RICHIE: And Lorien, quality?

DR. DALRYMPLE: Low.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Everyone voted

that is going to vote? Nineteen? Okay.

We have 11 voting moderate; 7 voting low.

And finally, consistency. One is high; 2 is moderate; 3 is low; 4 is insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

MS. RICHIE: Thank you.

CO-CHAIR CROOKS: Okay. I'm sorry. Two voted high; 16, moderate; 2, insufficient.

DR. PACE: Okay.

CO-CHAIR CROOKS: So, if we go with the majority, we --

DR. PACE: We would pass evidence.

CO-CHAIR CROOKS: -- would pass evidence.

DR. PACE: And just we will go back and check. Impact was passed and so was --

CO-CHAIR CROOKS: The performance was judged --

DR. PACE: Yes.

CO-CHAIR CROOKS: Performance gap was not demonstrated.

DR. PACE: Right. Performance gap was not demonstrated. So, technically, that would not meet our importance criterion. So, we will have a discussion here whether that -- you know, I think the importance or the opportunity for improvement reflected what was available to you. I guess the question is whether -- technically, this measure would stop here by not meeting that.

So, I am not asking you to change your vote, but if there is some rationale for moving forward?

DR. KLIGER: Well, if I can, as a non-pediatrician, it seems inconceivable to me that there is not a performance gap here. There is for adults. There has to be a substantial performance gap, even though there is no evidence for that. I think we should move forward, despite lack of performance gap

evidence.

DR. LATTS: I would agree, and it sounds like, while it is not in the submission, that the literature has shown that there are a substantial amount of kids who have hemoglobins under 10. So, one would assume, then, there are some that have hemoglobins over 10. Therefore, there is a performance gap.

CO-CHAIR CROOKS: Also, that will be the first job when they get the metric, is to find out if it being done or not.

DR. PACE: Right. So, we could ask PCPI to maybe provide us with some information from literature, like Rick mentioned, that shows what percentage of pediatric patients are anemic. You know, just as Alan said, there is probably something that you could do.

Okay. All right. So, any objections to continuing?

CO-CHAIR CROOKS: Any objections?

(No response.)

We're all comfortable with that? Okay.

DR. PACE: All right.

DR. BERNS: Can I just ask one question? I'm sorry, just things pop into my head.

In terms of the pediatric world for these patients, are there patients who would be excluded from this who would be under the care of a pediatric nephrologist but are above the age of 17, that we should just think about whether there ought to be some -- I don't know whether you can do it -- a revision in the age. Because, really, what you want to do is capture all of your patients who are under the care of a pediatric nephrologist who are on dialysis regardless of age, I would think.

DR. KASKEL: We all follow, depending on the institution and state, patients who are in the transition zone, getting them ready to go to you folks. In my center, we follow them until they turn 22, and

half our patients are over 17 where I am in the innercity, which may be a little unique. I don't know.

Michael, do you want to comment on that, too?

DR. SOMERS: I mean I think Barbara alluded to this earlier in her comments as well, that many of us have a proportion of our patients who are later adolescents and young adults. Some of the data that exists under the pediatric data include older adolescents within that. So, some of the health quality outcomes data, as well as some of the data looking in terms of detrimental physical effects of anemia, also include a fair number of older adolescents.

DR. KASKEL: I just want to mention, as far as this Best Pharmaceutical for Children Act, which is an act of Congress supporting research in pediatrics, the fact that this concept of anemia in CKD was chosen as a priority for 2012 was based on

performance data, that there is a gap in our knowledge for dosing, treatment, target values, prevention of the morbidity.

So, I don't know what message I didn't get across, but this was chosen amongst five or six areas of research from the NICHD for next year.

CO-CHAIR CROOKS: Okay. So, we can move on to --

DR. PACE: So, go on to reliability.

DR. LATTS: Karen, I wonder, given this additional discussion, should we take another vote on the performance gap or just --

CO-CHAIR CROOKS: Well, maybe you were out.

DR. LATTS: Okay. Sorry.

CO-CHAIR CROOKS: We offered anybody a chance to put up a counterargument or object, and nobody did.

DR. LATTS: Okay. Thanks.

CO-CHAIR CROOKS: So, we decided to

move on.

DR. LATTS: Right. I was just wondering, for the record, if we wanted to revote on the performance gap information. We were voting on what is in the submission.

CO-CHAIR CROOKS: It doesn't change the fact. There is insufficient evidence, but that doesn't mean there isn't one.

DR. LATTS: Right.

CO-CHAIR CROOKS: And so, we are going to go with common sense. In these things, we believe there is a gap. We think that the importance is such, and so on, that the Committee has decided to let it ride for now and move on to the next criteria. Okay.

DR. PACE: Rick, reliability?

CO-CHAIR CROOKS: Rick, would you like to --

DR. KASKEL: For reliability, we have, well, you can see the breakdown there.

DR. PACE: So, there was some difference of opinion.

Was reliability tested for the pediatric measure or is this also adult, the same?

DR. LATTS: This is adult, but the testing methodology we feel should hold true for a pediatric population.

DR. PACE: Other reviewers for this measure, any comments about reliability?

CO-CHAIR CROOKS: Alan?

DR. KLIGER: Just quickly, I don't personally see any reason why reliability testing for the adults should be any different than for the kids. So, in the specifications measures, I would suggest that we can accept the testing that has been done for adults.

DR. PACE: Okay. All right. Ready to vote on reliability?

DR. DALRYMPLE: This is Lorien.

Can I just mention the same concerns as before? In the e-specification there's a lot of unusual data elements that may not be relevant to the measure, and I

think some simple mistakes, like patient age, 18 and older, in the flowsheet, et cetera. So, that is in the attachment.

DR. PACE: Does the Steering Committee want to look at that? Do you want to take the same approach as last time? Is it kind of throughout again, Lorien?

DR. DALRYMPLE: Yes, and I think some things are just simple typos where at the top they say it is going to be patients age 17 years and younger, but then under IPP it says patient age 18 and older and then similar things with continuous veno-veno hemodiafiltration, multiple hemoglobins, including hemoglobin F and C, et cetera.

Now this one actually has exclusions. So, some of those would be thrown out. The hemoglobin S's, et cetera, would all fall under their exclusion criteria. So, they would be removed.

DR. PACE: So, what is the pleasure of the group? Do you want to continue on this

excluding the EHR specifications and get some response from the developer on those? Any objections to that or think it is unnecessary?

(No response.)

All right.

CO-CHAIR CROOKS: So, let's go on as Karen as outlined.

DR. PACE: Okay. So, we will get back to PCPI about the electronic specifications and the e-specifications for this one as well.

Okay. So, reliability. We can go ahead and vote, Tenee. High, moderate, low, insufficient?

MS. RICHIE: Lorien?

DR. DALRYMPLE: Low.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Nineteen, is that it? Okay, 20.

DR. PACE: All right.

CO-CHAIR CROOKS: Okay. One, high;

13, moderate; 4, low; 2, insufficient.

DR. PACE: Okay.

CO-CHAIR CROOKS: Validity testing. The same strategy, expert panel, the same expert panel. What do you know?

DR. PACE: All right. Okay. So, validity, of course, includes specifications consistent with the evidence, validity testing, which I believe Rick has said again face validity that was presented, I believe.

DR. KASKEL: Yes.

DR. PACE: Okay. And then, whether there is any issue with exclusions or not having risk adjustment.

Any discussion?

DR. DALRYMPLE: This is Lorien. Can I just clarify? The expert panel, this was the adult measure they were voting on? The results are identical. Or was this on a pediatric measure they voted on?

MS. CHRISTENSEN: This was on the pediatric measure.

DR. DALRYMPLE: Okay. Thank you.

DR. PACE: Okay. Any discussion?

(No response.)

All right. Let's go ahead and vote. Validity, high, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Has everyone voted who is going to vote? Okay.

All right. Nobody voted high; 16, moderate; 1, low; 2, insufficient.

So, we can assume that the next vote --

DR. PACE: Right.

CO-CHAIR CROOKS: -- would pass.

DR. PACE: Right. So, that would pass scientific acceptability. So, we need to see, were any disparities identified with this particular measure?

DR. KASKEL: There are disparities in some of the subcohorts and populations, yes. And there is data to show that African-American children with CKD and dialysis enter dialysis with lower hemoglobins, and there is data on that. Pediatric patients, yes.

DR. KLIGER: So, do the measure specifications scoring and analysis allow for identification of those subunits?

DR. KASKEL: As currently set up, no.

DR. PACE: And we are kind of working our way through. I should also preface this by saying we currently have a disparities project going where they are going to make some more recommendations about how to handle this in measurement.

But in the submission form, did PCPI talk about the disparities? Do you want to go to 2c?

Lauren, do you want to read?

MS. RICHIE: "The results of this

measure to be stratified by race, ethnicity, gender, and primary language, having included these variables as recommended data elements to be collected."

DR. PACE: Okay.

CO-CHAIR CROOKS: So, they have the data. And so, it is possible for them to put it out in a format where disparities can be analyzed.

DR. SOMERS: Correct. I was wrong. It's there.

DR. KLIGER: So, the measurement specs do allow for scoring and analysis by group? I'm just trying to understand that, because that is what we are being asked here.

DR. PACE: Right. So, do you want to --

MS. CHRISTENSEN: Sure. So, if you had a -- let's use the manual collection of a measure, just because it is easier to understand conceptually than an EHR or a claims. So, if you had a manual collection

form, you would simply indicate -- what did we say? -- race, ethnicity, gender, and primary language for that patient. Then, you would be able to run data analysis on those variables to group patients by different races, ethnicities, genders, or primary languages.

DR. KLIGER: So, the measurement specs have all of those in them right now?

MS. CHRISTENSEN: Yes. We recommend they be collected.

DR. PACE: Well, no, those aren't in the measure specifications currently. So, right now, we don't have -- I think probably the way to look at this, this is not going to make or break the measure going forward. So, if the answer is no, it's no. You don't have this stop at this point.

And certainly, when they bring the EHR specifications back, that can be noted, that those are specifically included. That should probably be also indicated in the kind of English language specifications.

Okay. So, shall we go ahead and vote on this? High, moderate, low, or insufficient.

MS. RICHIE: Lorien, disparities?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Twenty, the magic number.

We have 12 voting moderate; 1, low; 7, insufficient.

So, moving on to usability.

DR. PACE: Yes, right.

CO-CHAIR CROOKS: The next slide.

Rick, did you have any comments, or the Work Group, on usability?

DR. KASKEL: I think it is feasible; it can be measured, and it can be accumulated on a regular basis.

CO-CHAIR CROOKS: Yes, all raters rated it moderate or high.

DR. PACE: Okay. Any discussion?

(No response.)

All right. Usability, high, moderate, low, or insufficient.

Go ahead.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Stuck at 18. Anybody else voting? I guess 30 seconds is enough. Okay, let's stop it there.

Okay. Six rated it high; 14, moderate. So, it passes the usability criteria.

Then, feasibility.

DR. PACE: Right.

CO-CHAIR CROOKS: Rick, any comments on that?

DR. KASKEL: I think most have been high and moderate.

DR. PACE: Any questions or issues about feasibility?

(No response.)

All right, Tenee, let's start.

Feasibility, high, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Twenty, okay.

So, 12, high; 8, moderate. So, no problem with the feasibility criteria.

DR. PACE: Okay.

CO-CHAIR CROOKS: So, the final question is --

DR. PACE: Yes. Okay.

CO-CHAIR CROOKS: -- does this meet the NQF criteria for endorsement?

DR. PACE: Okay.

CO-CHAIR CROOKS: The clock is running.

DR. PACE: Right. Yes, no, abstain.

CO-CHAIR CROOKS: Yes, no, or abstain.

MS. RICHIE: Lorien?

DR. DALRYMPLE: No.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Has everyone voted? Okay.

So, 17 yes and 2 no.

DR. PACE: Okay. Very good.

CO-CHAIR CROOKS: Well, we're rolling now.

(Laughter.)

We're through anemia.

DR. PACE: Okay.

CO-CHAIR CROOKS: So, we are going to move to cardiovascular now without a pause, and we will first start by having those measure developers who have cardiovascular entries/submissions to please give us a brief description.

DR. PACE: Lauren, do you want to say who?

MS. RICHIE: I know PCPI is here. Do we have representatives from ActiveHealth Management?

MS. ALLEN: Yes.

MS. RICHIE: Okay. We'll let you go on the phone first. Thank you.

MS. ALLEN: Okay. Thank you.

So, this is Mureen Allen. I'm Senior Medical Director with ActiveHealth Management.

I would like to take the opportunity to thank the Committee and the NQF for giving us this opportunity to listen to the discussions about our measures and to contribute, where appropriate.

We have two measures that are up for review for the annual maintenance process, 0626, chronic kidney disease, lipid profile monitoring, which is a process measure, and 0627, chronic kidney disease with LDL greater than 130, which is an outcome measure.

There is a brief summary in your literature that was submitted with our form.

These measures address the gaps in care related to identifying patients with

chronic kidney disease who also have dyslipidemia and are at risk for ischemic vascular disease.

Our measures use clinically-enriched administrative data. More recently, we have also incorporated line items for Health Information Exchange data, data coming from electronic health records.

That's about all.

CO-CHAIR CROOKS: Okay. Thank you.

PCPI?

MS. AST: Thank you.

I'm Katherine Ast, Policy Analyst with the PCPI.

Our cardiovascular measures were originally created in 2007 with our Chronic Kidney Disease Work Group, and they have just been updated with our current Work Group with the updated evidence. They have been tested for reliability and validity. As well, they are currently in use with PQRS.

CO-CHAIR CROOKS: Okay. So, we

will start with 0626.

Ruben, you're up again.

DR. VELEZ: Good afternoon.

(Laughter.)

This is for endorsement maintenance. Essentially, it is lipid profile done on a time period of 12 months on anybody with CKD, essentially, a percentage of patients with chronic kidney disease from 1 to 6. So, it includes dialysis, and it includes transplantations. The denominator includes males over 10 years old, females over 13 years old, again, diagnosed with any stage of CKD.

There were only some general exclusions. There were no specific exclusions.

The Committee, the Work Group that worked on this measure, at least if we start looking at the impact, 1a, at the high impact, there was one intermediate, or insufficient -- I'm sorry -- one medium, one high, and two lows.

Comments?

And I would like to bring up that the next three measures we will be talking about have to do with lipids. There is one measure very close to this measure that we are going to talk later today. So, they are going to be very similar, except the denominator may be different, but that is just a comment.

DR. PACE: So, impact, variability in terms of the initial reviewers. Thoughts about that?

DR. NALLY: I was on the "L" side. My concern about this measure for CKD and others is that they rely on CPTs and physician diagnosis of chronic kidney disease, thereby potentially missing the majority of people that actually have chronic kidney disease who may be cared for by a primary care doctor for their diabetes hypertension and have a creatinine of one and a half without any recognition in the medical record that they have CKD. And therefore, they are basically

excluded from many of the measures we are going to talk about, be it cholesterol or blood pressure or other things.

I can tell you in our CKD registry at the Cleveland Clinic you can get into it either because you have too low GFRs or that the doctor has made a diagnosis with an ICD-9 code or has a listing in the problems. The overwhelming majority have inclusion into the registry -- and we are talking about over 60,000 patients -- have inclusion into the registry because of CKD diagnoses rather than the doctor making the diagnosis.

So, I don't have an answer for that in these different measures we will talk about, but I think the group needs to recognize that potentially we are missing probably a majority of patients with chronic kidney disease if we have as a threshold the doctor identifying them based upon a CPT or ICD-9 code.

DR. PACE: Okay. And maybe what we

can do is separate that. I think it is a very important part of specifications and validity. I think what we want to address right here is, if we had measures regarding lipid management, lipid monitoring, in this population --

DR. NALLY: Okay. So, I would be happy to proceed then.

DR. PACE: Okay.

DR. NALLY: My concern with this particular measure presentation is the review of the evidence is rather superficial and dated. Particularly, there are two randomized controlled trials in the dialysis population looking at cholesterols and statin use that are negative. Then, more recently, we have the SHARP trial which includes both CKD patients and dialysis patients, which is a positive trial. And in that sense, that information may actually strengthen the rationale for this trial, but it goes unmentioned.

DR. DALRYMPLE: And I agree. This is Lorien. I also scored it as low. I didn't feel like the data submitted -- it is higher, but based on general understanding of the field, I think you could have a view that it is moderate in impact.

CO-CHAIR CROOKS: Alan?

DR. KLIGER: I would suggest that lipids are a national health priority, and the issue around the strength of the evidence is something we can consider after this point. But, at this point, I think this question is pretty self-evident and we need to move past this one.

CO-CHAIR CROOKS: Thank you. Right.

So, are we ready to vote on impact?

DR. PACE: Okay. All right, Tenee?

High, moderate, low, insufficient. Impact is what you're voting on.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, we're up to 20. I think that is -- 21, okay.

Okay. We have 3 voting high; 12, moderate; 4, low; 2, insufficient.

So, next is the performance gap question.

DR. VELEZ: In the performance gap, they bring one study that showed that only 75 percent of patients with CKD had some type of cholesterol testing in a year. And they also bring some concerns about disparities of care in the population, whether commercial versus uninsured, whether diabetic hypertensives, and also race.

CO-CHAIR CROOKS: Okay.

DR. BERNS: I'm a little concerned that we are applying data on a lot of this from the wrong population, to this population. This is kids as well as adults. So, we are mixing two very disparate groups with data only as it relates to adults.

DR. VELEZ: That is completely correct, yes.

DR. FISCHER: I mean I think we will get to this, I guess, in the evidence, but to me there is a lot of heterogeneity. You have kids and adults. You have non-dialysis CKD and dialysis. Oh, and we also have prevalent CVD or cardiovascular disease and people without prevalent CVD, which means you are mixing primary and secondary prevention.

This seems very broad in scope, and a lot of the evidence and the importance may be different among any of those groups, dialysis/not dialysis, kids versus adults, or those with or without preexisting cardiovascular disease.

The way I read it, it seems like this covers all those groups. Or did they make accommodations that those will be treated separately?

CO-CHAIR CROOKS: Well, the issue

before us right now is performance gap. That may not -- you know, as opposed to a measure specification.

DR. PACE: And this is a monitoring measure. Do you think that --

DR. FISCHER: Only that I guess it depends on the performance gap, I guess the gap depends on if there is evidence that there should be a reason to be doing it. But that is the only reason why. I mean I think the performance gap and the evidence, I realize it is a discrete issue, but they are interrelated, right, to some extent?

DR. BERNS: I have another question. That is, to which physician population this pertains? So, is this nephrologists, pediatric nephrologists, adult nephrologists, pediatricians, family physicians?

DR. VELEZ: It's all of the above.

DR. BERNS: All of that? All of that?

DR. VELEZ: Yes.

DR. PACE: Developer PCPI, it would apply to any physician? I mean, I'm sorry. ActiveHealth?

MS. ALLEN: Right. So, this applies to any physicians who is taking care of a patient who has chronic kidney disease. So, it might cut across the nephrologists. So, there is a lot of feedback. So, nephrologists, if there was a primary care physician involved as well, that would also be measured as well.

DR. PACE: All right. So, Jeff, you brought up that the data on performance gap is about one particular group that is covered in the measure. Is that what your point is?

DR. BERNS: Yes, there is very limited data here about any performance gap, and what there is doesn't seem to apply to most or many of the patients to whom this might apply. And it doesn't really look at

all the different practice settings in which this might apply.

I am a little bit, well, I am more than a little bit uncomfortable with potentially endorsing or approving a performance measure that is going to apply to lots of different types of practices and types of physicians about which we are not experts.

MS. ALLEN: Could I just make one point? We did provide supplemental evidence with additional gaps-in-care studies. So, we did provide that last week.

DR. PACE: Okay. So, you said that was in the information you provided us last week? Okay.

MS. ALLEN: Yes. There is another study that looks at KDOQI hypertension, dyslipidemia and diabetes (telephonic interference) guidelines for CKD. It also talks about gaps in care in some of those, the measurement (telephonic interference). So, there are other studies.

DR. PACE: Okay. We are having a hard time understanding.

MS. ALLEN: I'm having a hard time, too. There is a lot of echo coming back my way. So, I have to apologize.

DR. PACE: Are you on a speaker phone or --

MS. ALLEN: Yes. Hold on a second.

DR. PACE: Can you pick up the --

MS. ALLEN: Is this better?

DR. PACE: Pick up the handset.

MS. ALLEN: Is this better?

CO-CHAIR CROOKS: It sounds better, yes.

DR. PACE: Yes, yes.

MS. ALLEN: Okay. So, I'll keep the handset up. Thank you.

DR. PACE: So, data on performance gap, you submitted more data? Is that what you were saying?

(No response.)

CO-CHAIR CROOKS: Oh-oh.

DR. PACE: She might have gotten cut off.

Okay. So, what we are talking about right now is performance gap/opportunity for improvement. Then, we will address the evidence and the specifications, if we end up moving forward with this.

So, right now, the question is, you know, is there information that supports that there is a performance gap on lipid monitoring? And the options are high, moderate, low, insufficient.

CO-CHAIR CROOKS: Let's vote.

DR. PACE: Okay. All right.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay. We have 8 Committee members voting moderate; 5, low; 7, insufficient. So, it is rather down to the -- I guess the mean would really be in the low category.

DR. PACE: Okay. We will continue.

Ruben?

DR. VELEZ: Does it stop there?

DR. PACE: No. Even though you have to have all three met, we will continue to do the evidence, so that we can look at importance all together and then address --

CO-CHAIR CROOKS: Right. We already let that go on one measure today.

DR. PACE: Right, right.

So, this is not an outcome. So, let's talk about the evidence. We'll talk about it all together. Then, we will rate quantity, quality, consistency.

So, Ruben?

DR. VELEZ: On the evidence, they mostly look at KDOQI. They look at 32 studies that I think KDOQI had. They mention about some tables and, essentially, talking about this lipidemia and CKD.

They do mention, and I quote, "Studies included are of mixed quality."

And they mention this. There are no RCTS testing the hypothesis that this lipidemia caused atherosclerotic disease in CKD.

So, that is what they bring in the evidence and the number and some of the quality discussion.

DR. PACE: And what is the evidence about? It is, obviously, not about monitoring. It must be about -- what is the --

DR. VELEZ: I mean, from what I see here, it is clinical practice guidelines that they are quoting, and they have selected some individual studies.

DR. PACE: And it is basically the link between CKD and hyperlipidemia?

DR. VELEZ: Correct.

DR. PACE: Other reviewers?

DR. BERNS: I'm not a reviewer, but I have a question.

DR. PACE: Oh, that's okay.

DR. BERNS: I guess, thinking about how this measure is set up, the question really is, I think, does monitoring lipids influence outcomes? Because there is no goal here. So, the question is, do you improve patient outcomes by monitoring? Is that a fair interpretation?

DR. PACE: Yes, and I think this is back in the category that we talked about with the first measure. It is about whether you assessed, and there are many steps that have to happen before you actually influence the intermediate outcome or health outcome.

And as someone mentioned, there is never going to be trials about how often you assess. So, it is always going to be indirect evidence, but, generally, from a performance measurement standpoint, the direction that we have been going is it is preferable to have something closer to the desired outcome. But there are circumstances, as we have talked about before, where there may be some

exceptions to that.

DR. VELEZ: And if I may remind the Committee, again, when we look at quantity, quality, and consistency, like it was well-stated, this is a measure that includes a lot of groups of people, of patients, a lot of subgroups.

CO-CHAIR CROOKS: Jerry?

DR. JACKSON: I am concerned about the observational study that links CKD to hyperlipidemia, and then the next study quotes reduction of mortality in CKD patients who are treated with a statin.

There are also a few, a small volume of observational studies showing that across the board statins reduce inflammation, micro-inflammation, and may have beneficial effects in CKD atherosclerosis. So, what is quoted here is a fairly loose association of benefit.

In other words, just the fact that patients have hyperlipidemia and are on a

statin and have reduction in mortality compared to CKD patients who are not on a statin with hyperlipidemia is not absolute proof of the benefit of this monitoring, I don't think.

DR. KLIGER: Can I ask the pediatricians in the room if they are aware of any evidence that lipid monitoring makes any difference to outcomes in children?

DR. KASKEL: We have data on the CKiD, the Chronic Kidney Disease in Children Study, recent data as of May, looking at 680-odd patients enrolled over the last 10 years at multiple time points in many of them, showing that, again, over half of them had lipid abnormalities. This is not dialysis. We do not have a large dataset to look at the dialysis.

Now whether the lipid abnormalities and the CKD is associated with adverse outcomes, that is what we are studying. And the number of the children in that group, the

teenagers who have abnormal blood pressures, et cetera, LVH, this is what we are looking at.

DR. KLIGER: And in your practices, are you using data from lipid monitoring to change what you are doing?

DR. SOMERS: I think more of us are, yes.

DR. PACE: Joe, would you repeat your comments about the evidence? Because you made some comments earlier about evidence.

DR. NALLY: Well, the general comment about this area was that there are at least two different bodies of evidence with randomized controlled trials that are not mentioned.

The first is in the dialysis population, the 4D trial, which are German diabetic dialysis patients, and the Aurora counterpart, both statin/placebo with negative trials. No specific difference in the ESRD population. A lot of spin as to why that may

have existed in terms of preexisting disease.

But the newest piece of information, called the SHARP trial, which is in 9,000 patients, 6,000 CKD before dialysis, 3,000 with dialysis, split between PD and hemo. Used lipid-lowering therapy that included a statin and showed statistically less cardiovascular events in that trial. No difference in renal outcomes.

But it was a positive trial that I think would at least bring a heightened awareness to the issue of dyslipidemias in CKD, whether or not they are truly causative in and of themselves, whether there are pleiotropic effects of statins or the other medication, to be announced.

But it, in my judgment, would at least strengthen the rationale for a measure, maybe not this measure, but a measure to be at least checking lipids, whether or not you do something with them. This is just a monitor question.

But this measure tends to be somewhat unfocused in the populations across pediatrics, the CKD, the dialysis, and transplant.

DR. PACE: Okay. So, we will start with quantity. High, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: For quantity, moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: One high; 9, moderate; 8, low; 3, insufficient evidence.

Next is the quality of the body of evidence. High, moderate, low, insufficient.

Is Lorien gone?

MS. RICHIE: Lorien, quality?

DR. DALRYMPLE: Quality, insufficient.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: I guess no one else is going to vote. All right. Oh, we're

up to 20? Okay.

Seven, moderate; 6, low; 7, insufficient evidence.

It does go down as a low at least.

DR. FISCHER: We are voting on what was presented in the application, correct, not things we know outside, but what was presented? I mean that was my understanding, Karen, from the beginning, was that we vote on what was presented in the application for the measure. Or I want to make sure I don't misunderstand.

DR. PACE: Right. Okay. What we talked about, that was our guidance for your preliminary evaluation.

One of the things that we want you to do as a Committee is, if people are aware of evidence, to bring that to the attention. Then, the Committee can use that, also, in their ratings in this meeting.

So, Joe mentioned a couple of additional studies.

CO-CHAIR CROOKS: Yes, you can take that into account.

DR. PACE: Right. And I think you have to have a discussion about that and have a discussion about how confident you are in terms of what we know about those studies that people bring up in the meeting or whether we need to ask the developer to go back and get something more. Okay?

DR. BERNS: You know, one issue that creeps up a number of times here, has crept up a number of times, is that the evidence may be good, so the SHARP trial was a well-done trial, but may not apply at all to many of the patients who would be included in this measure. So, we have to be very careful about thinking about both the quality and the appropriateness of that evidence to what we are discussing because sometimes I think there is a little bit of a disconnect.

DR. PACE: Right, and I think this is something we will need to figure out a

little bit better. Right now, that is kind of encompassed in quality, the directness of the evidence for what you are intending to measure. But perhaps we need to think about that.

Certainly, as you go through the rest of today and tomorrow, if you have some suggestions for us, we would definitely like to hear those.

DR. KASKEL: Just an aside -- and Michael is not here to support this -- but in pediatrics, we treat the lipid abnormalities not because they are symptomatic, but because we are concerned about what is going to happen to that patient when we transition them to you folks. And that is really the basis of our treatment. We don't have data to substantiate our treatment.

This is the truth. This is where we are with it.

CO-CHAIR CROOKS: Okay. So, based on Michael's point, do we need to revote the

last one on the quality of the evidence?

DR. PACE: Well, there was enough to continue, right?

CO-CHAIR CROOKS: Because I think we agree we could take into account newer information.

DR. FISCHER: Yes, I would just second the point that, once again, the way this measure is currently written, this is a wide swathe.

DR. VELEZ: And added to that are the comments we made already. I mean there are several good studies in the adult population that just came out, not in this huge group that were in this measure.

So, I think we need to keep thinking that this measure is a huge group of patients and we don't have data on them.

DR. PACE: So, going back to Peter's question, do you want to revote on quality or continue moving forward?

CO-CHAIR CROOKS: Does anybody want

to revote? No? Okay.

DR. PACE: Okay.

CO-CHAIR CROOKS: Let's go to consistency then. High, moderate, low, or insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Insufficient.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Four moderate; 5, low; 12, insufficient.

So, applying this to our grid, I think we get a no.

DR. PACE: Okay. The next slide, Tenee.

So, basically, with low on --

CO-CHAIR CROOKS: Quality and consistency.

DR. PACE: -- on consistency, it would mean it doesn't pass evidence. We also had a problem with performance gap, I believe.

CO-CHAIR CROOKS: Consistency was

really more insufficient rather than low.

DR. PACE: Right.

CO-CHAIR CROOKS: But maybe that needs to be taken into account in the grid, if you are going to let people vote that way.

But I think we can agree it hasn't really, for several reasons, it hasn't really met the criteria for quality or consistency.

DR. PACE: So, the next question, then, on evidence, the question would be again, does this measure merit consideration for an exception to the evidence?

Yes, Lisa?

DR. LATTS: I guess what I am struggling with is that this is a very important measure, but it is the mishmushing of all the different kidney patients together. If they divided it up into 3s and 4s and then ESRD and then transplant, I would feel much more comfortable.

DR. KASKEL: There is a competing measure that we will be talking about,

correct?

DR. PACE: So, just to clarify, I mean, going forward, there's a couple of things that could be considered. We will continue to finish up this measure. As you know, you can make some recommendations to the developer. If you think it is really an important measure, but that stratifying or restricting the denominator to where the evidence leads, that would be one option.

One option would be to just vote on this measure as is, and then we can come back to it to look at after we have looked at all the lipid measures.

DR. LATTS: If we did ask them to split it up, would that, then, have to wait until the next review cycle, whenever that would be, or could it be --

DR. PACE: I don't think it would necessarily have to wait until the next review cycle. I mean we would have to have a discussion with the measure developer.

ActiveHealth, are you still on?

MS. ALLEN: Yes, I am.

DR. PACE: Okay. I think we lost you a little bit before.

Should the Committee decide that they want you to limit the denominator or do some stratification, is that something that could be accomplished?

MS. ALLEN: Certainly. We can certainly break the groups into pediatrics versus adults. And then, we can also break it out versus the different stages of kidney disease. We can certainly do that.

DR. PACE: All right.

CO-CHAIR CROOKS: Yes, are we asking, though, just to stratify the different groups? So, now we have one for ped, now we have one for post-transplant, now we have one for -- or to just pick a group where there is the most evidence that the metric would more likely pass?

MS. ALLEN: Okay.

CO-CHAIR CROOKS: Ruben?

DR. VELEZ: My recommendation would be to vote with what we have now, and like you suggested, Karen, we could come back after we see all the other measures we are going to discuss.

But I have trouble when we start -- too many cooks in the kitchen, you know, can be a problem.

CO-CHAIR CROOKS: Wise words.

Okay. And also, I would just remind the group that we are only going to consider this metric on its own first. We are not going to try to say there is a better one coming down or a worse one coming down the pike. Okay.

DR. PACE: So, is there any thought that you want to move forward? It sounds like there is still an issue with the evidence focused on being consistent with how the measure is specified.

Is there anyone who wants to invoke

the exception for expert opinion, since this is an assessment measure?

(No response.)

Okay. All right.

CO-CHAIR CROOKS: I see no hands.

DR. PACE: Okay. So, basically, this measure would not pass importance to measure and report. But, as we kind of get through the sets of measures, we can certainly come back to this, pick it up again, if the Committee desires. Okay.

CO-CHAIR CROOKS: Okay. So, we will move to 0627, chronic kidney disease with LDL greater than or equal to 130, use of lipid-lowering agent.

Dr. Nally?

DR. NALLY: Thank you.

This is a renewal submitted by the same ActiveHealth Management group. The description of the measures is "the percentage of patients with chronic kidney disease and an LDL greater than or equal to 130 that have a

current refill for a lipid-lowering agent".

The numerator is the patients with current refill for a lipid-lowering agent. The denominator, all patients age 18 and older diagnosed with CKD, including CKD 5, dialysis, or transplant, and an LDL above 130.

Now, as I read that one time, I read it as all patients with CKD, including 5, dialysis, transplant. And one could interpret it that way because in their definition there are some general codes for CKD; whereas, there are more specific codes for CKD 5. So, that is one confusion I have right out of the start. Then, there are some general exclusions.

Is it possible to ask the measure steward at this time, are they trying to limit it to CKD 5, dialysis, and transplant, and not have codes in there with the general CKD or general nephrotic syndrome, et cetera?

PARTICIPANT: On page 7, they show it as being CKD 5.

DR. NALLY: But, then, they include CPT codes for 585 NOS and nephrotic syndrome, and other things. So, I want to be sure that they are, indeed, limiting it to CKD 5, dialysis, and transplant.

Is that correct, Measure Steward?

MS. ALLEN: That is correct. Where you see the NOS code, that is in conjunction with a creatinine clearance between 0.1 and 14. So, it is not an NOS code by itself. It is in conjunction with a creatinine clearance.

DR. NALLY: Okay. Thank you for that clarification.

MS. ALLEN: You're welcome.

DR. PACE: So, any comments about impact that you want to make? Then, we will vote on that and then go on to the other.

DR. NALLY: Well, I think, as was just articulated, the issue of dyslipidemia, cardiovascular disease in the CKD and dialysis population is, indeed, important. However, again, the measure per se, the review of the

evidence is, again, somewhat cursory. It doesn't bring into context some of the trials that we just talked about, et cetera.

DR. PACE: Okay.

DR. NALLY: So, are there other comments of the other reviewers?

(No response.)

DR. PACE: Let's go ahead and vote, then, on impact. Then, we will move on to performance gap and evidence.

Okay. So, impact, high, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

DR. PACE: Is everyone done? Okay.

CO-CHAIR CROOKS: The results are 3, high; 14, moderate; 3, low. Nicely balanced. Okay.

Next, performance gap.

DR. NALLY: The performance gap cites a single study recently published this

year, primary care practices regarding a moderately-large number of patients. But we don't have a great deal of information about the quality of that evidence. They do suggest that there is a performance gap.

CO-CHAIR CROOKS: Joe, as you pointed out, as you clarified, the denominator is the Stage 5 in dialysis patients. This performance gap data is Stage 3 and Stage 4.

DR. NALLY: Correct. There is a dichotomy there. Once they have clarified that this is 5 in dialysis and transplant, the dichotomy exists that the single study cited is not applicable to that patient population that the measure addresses.

CO-CHAIR CROOKS: Ruben?

DR. VELEZ: And again, I am somewhat confused. I understand the comment that was made. But when we look at the numerator, especially the denominator, it includes all the CKDs and nephrotic syndrome.

DR. NALLY: So, I specifically

asked the measure steward that question, and some of the specific codes that are probably related to nephrotic syndrome do not apply except if they have CKD 5. All the rest of the codes are dialysis, ESRD, and CKD 5 transplant codes.

Additional comments or questions?

(No response.)

So, it is really not possible to say that there is a performance gap when the evidence is disparate.

CO-CHAIR CROOKS: Could the developer, do you care to defend that, the way you presented the performance gap?

MS. ALLEN: If you can give me a couple of minutes? I just need to bring up the actual description of what we put there.

CO-CHAIR CROOKS: We are not hearing you very well.

MS. ALLEN: I just need to bring up the actual measure, just to see the description of what we put there.

But the performance gap is really based upon the recommendations that were made by KDOQI in terms of screening patients with Stage 5 disease for dyslipidemia and then treating those patients who have an abnormal LDL. So, this is part of the general recommendations made by the guideline.

DR. PACE: This measure was previously endorsed. And I can't remember if we asked or if you provided -- I'm trying to pull it up -- performance gap information on the actual measure.

Joe?

DR. NALLY: Three weeks ago, I had a telephone conversation with Lauren noting that this was kind of a private entity submitting the measure, but I couldn't find anywhere publicly reported kind of the performance of this measure. And Lauren was going to address that question to the measure developer, and at least I don't have access to that information.

Is there publicly-available information on the last year and a half or two of outcomes of this measure?

MS. ALLEN: We don't report out publicly the performance of the measures. Typically, what we will do, on behalf of clients, we will generate reports that we give to our clients specifically.

What we anticipate is this year, and probably next year, as we participate in a hospital care initiative, that a lot of our measures will be reported publicly at that point in time through our clients. But we don't directly report out publicly on measures.

DR. NALLY: Unfortunately, that makes it very difficult for us to look at reliability, usability, feasibility, and performance gap, if we don't know how the measure has gone.

MS. ALLEN: We can certainly give you the measures based upon our book of

business and based upon some of the testing that we have done. But in terms of actually reporting publicly, that, again, as I said, is usually done through our clients.

DR. PACE: Okay.

DR. NALLY: Thank you.

DR. PACE: All right. Okay. So, any other comments about performance?

CO-CHAIR CROOKS: Is it possible that we could put this on hold, so to speak, and let them go back and look at their data for performance gap and report to us at least where they are? They have this a year and a half or two years. I think if we are being asked to re-endorse it and they have data, we should see the data.

MS. ALLEN: We'll be happy to provide that for you, if you give us the opportunity.

CO-CHAIR CROOKS: Alan?

DR. KLIGER: I guess from a systems standpoint, it is tough to do that because we

would apply that same principle every time we find No. 4, insufficient evidence. We would always say, well, let's table it until they give us better evidence.

My urging would be for us to vote today on all the measures with the data that we have and then to move out from there.

CO-CHAIR CROOKS: But what is different here is that we know they have, she is saying they have that information. And so, it is just sitting there. They should have supplied it. The fact they didn't supply it with the application, is there a problem, I guess, when you can say they should have and they didn't, and they missed their chance?

DR. PACE: Well, let me ask, because we specifically asked for information on the opportunity for improvement. We went back to you, and you just gave us your original response. So, I guess that is a question of whether you can provide performance on the measure as it is specified

based on the data you have over the past year and a half. Even if you are not going to identify individual physicians or facilities, you know, distribution of the scores by quartile or decile, however, that is a question. It is unclear, because you just reiterated your original response.

MS. ALLEN: Is that a comment for me? Hello.

CO-CHAIR CROOKS: Yes, we hear you.

MS. ALLEN: I'm sorry. Yes?

CO-CHAIR CROOKS: Are you using the handset now?

MS. ALLEN: Yes. No, no. Yes, I am using the handset. I'm sorry. Yes.

I think part of the problem that we have, and certainly the NQF staff knows that we have had some difficulty in terms of timing. However, if you wish to get our book-of-business numbers in terms of the performance of the measure, we can actually give you that. What we have given you are

numbers that we had based upon a client or two that we had, and then based upon some testing on real data. But, then, we can certainly give you numbers for our book of business, so that you will have that to evaluate the measure.

DR. NALLY: I still have concern that, even if this data would be forthcoming to this group, it is still not being publicly reported, which would be the mission to establish good health for the country.

DR. PACE: But that's also the case with the PCPI measures at this point. It is being reported, but it is not publicly available for anyone to view the performance data.

So, that is definitely a goal of NQF. And if it is not, we want some plans to move towards that direction. And certainly, reporting it is at one spectrum of the accountability or transparency scale. So, it is moving in that direction.

And so, I understand your question. We would definitely get to that under usability.

So, Helen, do you have any suggestions on what we should do?

DR. BURSTIN: My general feeling is, if there is information that you can gather in a timely manner, you should have the full information when you make the assessment.

CO-CHAIR CROOKS: All right. So, unless there are objections, then I think we will just go ahead and vote based on the information they supplied.

Mike?

DR. PACE: Okay. Michael, did you have a question.

So, I'm sorry, I don't know your name that's on the phone for ActiveHealth.

MS. ALLEN: This is Mureen Allen.

DR. PACE: Mureen, do you, by any chance, have that data available, I mean that you could tell us now? I mean, do you have

any information that you are looking at about your measure that you could verbally tell us?

MS. ALLEN: I don't have that right now.

DR. PACE: Okay.

MS. ALLEN: I would have to go back to our team that does the data analysis. So, I don't have that available right now.

DR. PACE: So, perhaps what we could do is vote on it, based on what we know. And again, maybe we will have to think about, after we get through this list, if there are some opportunities to provide some information that would perhaps change the course of how things have gone, that we will relook at those. Does that make sense?

CO-CHAIR CROOKS: Right. No guarantees it will change anything, but we are willing to look at it.

DR. PACE: Okay.

CO-CHAIR CROOKS: If it submitted in a timely manner.

DR. PACE: Okay. So, importance to measure and report, is there a performance gap?

I guess I will also mention, if the Committee is aware of evidence about, again, even though it is not specific for this measure, the measure about the less than 130, so it is an intermediate clinical outcome. So, we can certainly hear from the Committee if you have knowledge about evidence about opportunity for improvement or performance gap. Okay.

CO-CHAIR CROOKS: I don't see any forthcoming.

DR. PACE: All right.

CO-CHAIR CROOKS: So, we are going to vote, 1b, performance gap, high, moderate, low, or insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Insufficient.

MS. RICHIE: Insufficient?

DR. DALRYMPLE: Correct.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Well, the insufficient carries the vote with 16, and 1 low; 3, moderate.

DR. PACE: Okay. So, we will follow our process. We will still ask you to evaluate evidence because that also may determine whether you would even want more information on opportunity for improvement.

So, let's see, who was presenting on this one? Joe?

CO-CHAIR CROOKS: Joe.

DR. PACE: Joe, do you want to talk about the quantity, quality, and consistency of the evidence for this same measure? So, we will finish out importance and then see where we're at.

DR. NALLY: I believe we have already made some of those statements about the evidence already in terms of the measure as submitted with small numbers of studies

cited in a dated fashion without any randomized controlled trial information.

DR. FISCHER: I had one question. This SHARP study, do you know what the achieved LDL was in the SHARP study? Because that actually, I mean it just appears this was the LDL over 130.

DR. NALLY: Unfortunately, everything was reported in international units. The cholesterols were like 5.3, which means times 40. So, they are about 230 on the way in with an LDL of about 120, I think.

We should look that up. My secretary is on it. Give me a minute.

(Laughter.)

DR. PACE: Okay. Any other comments on the evidence from the reviewers, the preliminary reviewers or additional Committee Member comments?

(No response.)

Okay. Well, then, let's go ahead and vote on quantity of studies for evidence

for this Measure 0627. High, moderate, low, insufficient.

MS. RICHIE: Lorien, quantity?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, let's go with that.

Six moderate; 9, low; 5, insufficient. So, that will come out really as a low.

DR. PACE: All right.

CO-CHAIR CROOKS: Okay. The next is the quality.

DR. PACE: Quality of the body of evidence.

Go ahead and start it.

High, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Twenty, okay.

Four voted moderate; 11, low; 5,

insufficient.

And finally, consistency. High, moderate, low, or insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: For consistency, low.

MS. RICHIE: I'm sorry, was that low?

DR. DALRYMPLE: Correct. Low.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay. Three moderate; 10, low; 7, insufficient.

I think we would have to rate all three categories as low or worse.

DR. PACE: Okay.

CO-CHAIR CROOKS: And so, that would yield a no from the diagram.

DR. PACE: Right, right.

Okay. All right. So, we are --

CO-CHAIR CROOKS: So, to sum up importance then --

DR. PACE: It would not meet importance.

CO-CHAIR CROOKS: It did not make the performance gap or the body of evidence.

DR. PACE: Right. Okay.

CO-CHAIR CROOKS: So, we're rolling.

Okay. So, we're up to 3:30. What do you think?

DR. PACE: Do you want to take a break?

CO-CHAIR CROOKS: Well, how are we doing out there? Are you getting your second wind like me?

(Laughter.)

You can tell, can't you?

All right. Well, let's do one more then now.

DR. PACE: Okay.

CO-CHAIR CROOKS: 1668, laboratory testing. Joe has this one also.

DR. NALLY: Thank you.

So, we are in a similar vein here. But, as contrast the other two, this is a new measure submitted by the AMA PCPI team.

The descriptor is "percentage of patients age 18 and older with a diagnosis of CKD Stage 3, 4, or 5, not receiving renal replacement therapy, who had a fasting lipid profile performed and results documented at least once during the past 12 months".

So, the numerator is the patients who had the fasting lipid profile performed an documented. The nominator are all patients age 18 or older with CKD 3, 4, 5, not receiving renal replacement therapy.

So, this is, given our other discussions, restricted to CKD 3, 4, and 5, not dialysis and transplant, age 18 and older. So, we have limited it to adults and non-dialysis CKD.

DR. PACE: So, why don't we go ahead and vote on impact, and then we can move on to performance gap?

Impact, high, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. NALLY: I think the group as a whole had either moderate or high in terms of the impact, if I read that correctly.

DR. PACE: Okay. Right.

DR. NALLY: Are there comments from the group?

(No response.)

DR. PACE: So, we will stop it and restart this. Sorry.

DR. NALLY: Well, my bad.

DR. PACE: No, that's okay.

CO-CHAIR CROOKS: No, you were right to ask for other --

DR. PACE: You're right, yes.

Okay.

DR. NALLY: Can you say that again, Peter?

(Laughter.)

CO-CHAIR CROOKS: As usual, Dr.

Nally is correct.

(Laughter.)

Does that make you feel better?

DR. NALLY: I'm leaving.

CO-CHAIR CROOKS: Okay.

DR. NALLY: So, do any of the group have specific comments?

(No response.)

Again, I would also observe that I don't believe those randomized controlled trials were included in this body of evidence, either, up to the SHARP trial.

DR. PACE: Okay. Well, let's vote on impact, and then we will get on to the --

DR. NALLY: Thank you.

DR. PACE: Okay. Tenee?

This is for impact. High, moderate, low, insufficient.

MS. RICHIE: And Lorien, impact?

DR. DALRYMPLE: Moderate.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Twenty-one.

Six voted high; 14, moderate; 1, low.

Okay. Is there a performance gap?

Joe?

DR. NALLY: In this case, we are provided some information about demonstrating a performance gap, including the similar data from the 2008 where about 56 percent of patients did not receive the optimal care, which would be getting a lipid profile.

And there are three or four studies cited related to performance gap from USRDS and other sources.

CO-CHAIR CROOKS: Okay. More discussion on this?

(No response.)

Okay. Let's vote. 1b, performance gap, high, moderate, low, or insufficient.

MS. RICHIE: Lorien, performance gap?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay. I think we could stop there.

Of those voting, 5 voted high; 13, moderate; 1, insufficient.

Okay. So, on to --

DR. PACE: So, this is a process measure. We will move on to talk about the evidence. Quantity, quality, and consistency. This is basically an assessment measure. So, it is not proximal; it is more distal to the desired outcomes.

And so, Joe, you were going to make some comments about the evidence.

DR. NALLY: The evidence here in a way reflects some of the discussions we have had. But, in addition, there is considerable more text or meat to the discussion related to performance gap, to disparities, and other lines of evidence, with citations related to disparities and other issues.

DR. PACE: Okay. And we have just

talked about opportunity for improvement. So, now we are on to the clinical evidence, 1c, right?

Did you guys vote on this? Am I wrong?

CO-CHAIR CROOKS: No, you're right.

DR. PACE: Okay.

DR. NALLY: I'm sorry, did I confuse the issue?

CO-CHAIR CROOKS: Well, according to the spreadsheet, the group -- let's see --

DR. PACE: Right. So, we had you vote on opportunity for improvement and performance gap. Do we need to go back to talk about that? No? Okay.

So, now we are talking about 1c, the evidence, the quantity, quality, and consistency of the body of evidence.

And who else reviewed this measure? It's 06 -- no, this is 1668.

DR. DALRYMPLE: This is Lorien. I was one of the reviewers.

I did initially -- I'm sorry, can you hear me, Karen?

CO-CHAIR CROOKS: Now we hear you.

DR. DALRYMPLE: Okay. I was also one of the reviewers, and I did put insufficient initially for quality and consistency. That was primarily based on what was (telephonic interference). I would modify that based on this (telephonic interference) so far regarding this body of literature.

DR. NALLY: I mean, in brief, they basically quoted KDOQI guidelines in this area, which is a B level of evidence, and did not bring into play the randomized controlled trials that we mentioned previously.

DR. KLIGER: So, then, just to quote what they said, they started with 258 trials that they reviewed, but they did not give us the number that they eventually came out with after their filters. So, in terms of quantity, we have no data here that will answer that question.

DR. PACE: This is about assessing lipid monitoring. What is the evidence about? Is it about treatment or is it about association of lipid levels to complications?

CO-CHAIR CROOKS: They say the principal reason to evaluate dyslipidemias in patients with CKD is to detect normalities that may be treated to reduce the incidence of ACVD, which is unproven, but that is their assertion here.

DR. NALLY: And, then, they also open the question about progression of CKD and with dyslipidemias and/or their treatment. But they are speculative, I guess, at least prior to SHARP.

CO-CHAIR CROOKS: SHARP isn't quoted in here.

DR. NALLY: No, that is, again, missing from this measure also.

DR. KLIGER: Again, just in terms of the question that we are being asked in terms of numbers, if you go through what they

went through, they looked at lots of studies, it is really clear. They don't give us the final number, including manuals that were added in, et cetera. But we are talking in terms of hundreds, not in terms of two or four or eight. I am just quoting from their form.

DR. PACE: Right. Okay.

CO-CHAIR CROOKS: Yes. So, I think the quantity question may be the easier of them.

DR. PACE: Okay.

CO-CHAIR CROOKS: We could probably vote now.

Okay. Let's vote the quantity question. High, moderate, low, insufficient.

MS. RICHIE: Lorien, quantity?

DR. DALRYMPLE: High.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: So, 13, high; 6, moderate; 1, insufficient.

The next question is the quality of the body of evidence. I think it is clear

that they are not addressing directly what the metric is, which is that the measurement is done.

DR. PACE: Discussion about the quality of the body of evidence?

(No response.)

And, Joe, you're saying that it is missing the latest --

DR. NALLY: Well, it is the same comment applied broadly across the board. A lot of this is inferring what is good for the general population may be good for the CKD population who have an increased prevalence of heart disease. The proof of concept, I guess, would be the SHARP trial. And either because of timing or whatever, that is not included in the presentation.

DR. PACE: But you're saying the SHARP trial did support the hypothesis?

CO-CHAIR CROOKS: Is this the same population as the SHARP trial, CKD 3, 4, and 5, not on dialysis?

DR. NALLY: Well, of the 9,000 people in SHARP, 6,000 were pre-dialysis CKD, 3,000 were dialysis. Now, as it turns out, over the course of the trial about two of those six thousand ended up coming to dialysis but were analyzed under intention to treat. But it is clearly a 9,000-patient study with CKD, the majority of which were non-dialysis.

CO-CHAIR CROOKS: Did it include CKD 3 as well, the SHARP? Because that is the largest group of CKD by far.

DR. NALLY: The average GFR in SHARP as baseline was 27. They have it displayed out on an introductory table in terms of CKD 3s, 4s, and 5s. But the short answer is, yes, CKD 3 was included, and it included, I think, about 29 percent of the non-dialysis group or CKD 3.

That is by making some slides in the last week. But if certain people over there with his computer would, you know --

(Laughter.)

CO-CHAIR CROOKS: Alan?

DR. KLIGER: The other thing, again, they quote that both the NKF Task Force and the KDOQI Work Group that looked at these data, they talk about CKD, and they don't specify which levels, but CKD. Both had strongly endorsed measures. So, some of their claims are piggybacking on those two groups

that have done this work before.

CO-CHAIR CROOKS: But the point I want to make is, if this is the same population as the SHARP trial, even though it is not brought up there, and there is evidence that treating does reduce cardiovascular disease, improves cardiovascular outcomes, then, in my mind, that would improve the quality of the body of evidence, that, therefore, screening is worthwhile because it leads to treatment that makes a difference.

DR. NALLY: Correct.

DR. BERNS: SHARP eligibility was men with creatinine over 1.7 or above and women 1.5 or above. So, at least some of them would have had early stages of CKD.

So far, I am only able to find that they reported percent reduction in lipids rather than --

CO-CHAIR CROOKS: Oh, really?

DR. BERNS: -- achieved LDL. Their average or their mean eGFR baseline was 27

amongst the 6,247 non-dialysis patients, with a standard deviation of 13. So, they would have been mostly Stage 3 and 4.

CO-CHAIR CROOKS: So, Jeff --

DR. NALLY: Ruben, do you have the paper in front of you?

DR. VELEZ: No. For the Aurora and the SHARP, the SHARP is really the only so far that has shown with treatment improvement in cardiovascular mortality in dialysis patients. All the three studies showed improvement in the CKD world with treatment. But the other two did not show improvement in the ESRD world. So, that is the difference with the SHARP.

DR. NALLY: But since this is limited to CKD 3, 4, and 5, this would, again, strengthen the evidence for this particular measure.

CO-CHAIR CROOKS: So, any other comments before we vote?

(No response.)

Okay. Let's vote on the quality of body of evidence. High, moderate, low, or insufficient.

MS. RICHIE: Lorien, quality?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: One high; 19, moderate; 1, low.

Okay. So, let's move on to consistency. Any discussion here?

(No response.)

Shall we vote? Okay.

Consistency results across the body evidence, high, moderate, low, or insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay. Sixteen moderate; 4, low; 1, insufficient.

So, our rating of the body of evidence would give it a pass then, I think.

DR. PACE: Right.

CO-CHAIR CROOKS: Okay. So, do we move to the next --

DR. PACE: Right. So, and they passed on impact and importance opportunity and evidence. So, we will go on to reliability.

CO-CHAIR CROOKS: So, we don't need to vote on this, right? Everybody is comfortable with a yes?

DR. PACE: Right. It just is a default.

CO-CHAIR CROOKS: It has to be.

DR. PACE: Right.

CO-CHAIR CROOKS: Can't change it. Sorry.

DR. PACE: Right.

CO-CHAIR CROOKS: Okay. All right. Acceptability, measure properties, reliability.

Joe, are you still up? Yes, he's still up.

DR. NALLY: Again, we are presented some evidence on these issues, including then some data samples from four nephrology practices with basically 30 CKD patients per practice with kappa statistics that seem fairly reasonable.

My concern is that they have selected nephrology practices for these people with CKD whereby the nephrologist is likely to use some type of CPT or ICDM diagnostic code, and that, again, we may be missing large numbers of people in the population that are cared by the non-nephrology physicians of the world in terms of internists, primary care doctors, et cetera. So, we are given data, but limited to nephrology practices.

CO-CHAIR CROOKS: So, you are addressing really the specifications, which comes, also, in the reliability consideration, I guess. That their specifications will leave out large numbers of patients. And while I agree with you, I don't know we can do about

it.

DR. NALLY: Like I said before, I don't know the --

CO-CHAIR CROOKS: In Cleveland Clinic and Kaiser, we have systems.

DR. NALLY: Right.

CO-CHAIR CROOKS: And anybody who walks in our door, we know basically whether they have CKD or not who joins the health plan, and the same for the Cleveland Clinic. But for most of the country, that is not an option.

DR. NALLY: Correct.

CO-CHAIR CROOKS: And so, I think we have to decide, is this better than nothing?

DR. WELCH: Is that a function of the measure or the function of --

CO-CHAIR CROOKS: Turn on your microphone and say that again, Janet.

DR. WELCH: Is that a function of the measure itself or a function of the

process? That is the question I am asking myself. Because if it is the measure, then you would have to go back and fix it. If it is to try to think about, well, how can we be more inclusive, that is really not a function of the measure itself.

CO-CHAIR CROOKS: Right. So, if I can rephrase what you are saying then, we have to kind of take it as it is; I think we have to take what is given to us and judge it on its merits. It is unfortunate that it doesn't have a broader, can't include everybody, but is that your point?

DR. WELCH: I think so, yes.

CO-CHAIR CROOKS: Okay.

DR. PACE: Okay. I think the developer had a clarification.

CO-CHAIR CROOKS: Clarification, please.

MS. AST: All right. Thanks.

We have been continuing to work on this measure set, just for your information.

And if it is helpful to this measure, we decided to specify a different measure that we weren't able to submit because we didn't have testing data. But to define the diagnosis of CKD, it can be identified in one of two ways, a diagnosis of CKD Stage 3, 4, or 5 -- this is a different measure now -- CKD NOS, or two eGFR lab results of less than 60 more than 90 days apart.

So, we are aware of this issue, and we wanted to capture just what he is saying in a different measure. And I believe -- I mean we would have to talk with the Work Group about it -- but we could discuss doing the same thing for this measure, if it was appropriate.

And just a further clarification, non-nephrologists can use this measure. You know, that is not the problem. But I do understand you are talking about capturing the patients without the --

DR. NALLY: Correct. An internist

can write 585.4 down. The question is, do they? Or particularly Stage 3, do they?

DR. LATTS: I think they do.

DR. NALLY: Not in our 66,000 people they don't. A lot less frequently than you think.

But here's the other question I didn't think about until you brought this up. It is you are talking about percent of CKD patients. Is this per practice, per doctor? I mean, who is on the receiving end of this, only nephrologists, every internist in the community, et cetera?

DR. LATTS: Are you asking about the testing?

DR. NALLY: No, no.

DR. LATTS: I'm sorry.

DR. NALLY: The evaluation process is a percent CKD patients, right, who have this maneuver done? Is that per Dr. Berns or per the University of Pennsylvania or the City of Philadelphia? I mean, who is being

critiqued?

MS. AST: It's a physician-level measure.

DR. NALLY: Thank you.

CO-CHAIR CROOKS: Question for the developer regarding reliability testing again: was this element testing? In other words, you had more than one reviewer look at a given patient's data to see if they extract the same information?

MS. CHRISTENSEN: Yes, this is a part of the same. It is all one study that was conducted.

CO-CHAIR CROOKS: Okay. Thank you.

Okay. Other comments, questions on specifications and reliability testing?

DR. PACE: And Lorien, are you still there?

DR. DALRYMPLE: Yes, I am.

DR. PACE: I know you're the one who delved into the electronic specifications. Did this measure have electronic

specifications, e-specifications as well?

DR. DALRYMPLE: This one does. I think my primary question, it looks like I wrote down when I reviewed them, is this CPT II codes or actual lab results, or both, would be used to ascertain the numerator? That wasn't clear to me.

MS. AST: We received word before we came, also, that our specifications team has neglected to include the lab, the link codes. So, those have now been updated, but you have not seen the new specs.

So, we apologize that those weren't included originally. So, it is meant to be both.

DR. DALRYMPLE: Okay. So, depending on who implements it, they may either use CPT II codes or they may actually pull direct lab data?

MS. AST: Correct.

DR. DALRYMPLE: Okay. Thank you.

CO-CHAIR CROOKS: Okay. So, I

think we're ready to vote. Reliability, high, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Reliability, moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Results: 2, high; 14, moderate; 4, low.

Okay. On to validity.

DR. PACE: All right. And was this validity face validity again for this measure, I believe?

CO-CHAIR CROOKS: That's what it says.

DR. PACE: Okay.

DR. NALLY: I believe so.

DR. PACE: All right.

CO-CHAIR CROOKS: Expert panel voting.

DR. PACE: And were there any issues with threats to validity, with exclusions? Risk adjustment wouldn't apply?

Yes?

DR. FISCHER: I know this has come up before, and I think this will come up again. I just want to make sure I understand how the denominator is being defined. So, it is either laboratory criterion for low eFGR or ICD-9 codes?

MS. AST: Yes, from what I understand from our specifications team, depending on whether it is an EHR or claims or paper.

DR. FISCHER: There have been two papers that have shown that ICD-9 codes that identify CKD overall -- we're not even talking about 3 versus 4 versus 5 versus 2 -- you know, there is reasonable specificity, but the sensitivity is not that great, meaning you are going to miss a lot of people, and there are inaccuracies.

But one study was from Medicare claims data. The other was from, I think, from VA and Medicare data. I don't know what

it is like in other areas. It may be different based on the incentive of physicians to code based on billing. The VA doesn't bill, et cetera.

But that is the type of validity data that I think would be a little bit interesting to think about because the papers I am aware of, two of them, there were problems in that, once again, you have some specificity, but not a heck of a lot of sensitivity. And therefore, you would miss people.

MS. AST: I apologize. I think I misspoke. I was talking about the numerator. I'm sorry, I misspoke about that.

At this point, the denominator would simply be the codes, the ICD-9 codes.

DR. FISCHER: Okay.

DR. NALLY: Which will tend to clearly underreport true CKD as it exists in the wild.

DR. FISCHER: Well, and the other

thing I would just mention is I would also be concerned. If we are really trying to target 3, 4, and 5, I think you also -- and I am not aware of anyone who has ever evaluated this -- that a CKD code, people might use that for someone, right, who has intact GFR? And maybe that is okay, but you maybe have people who have really 1 and 2, based on consensus definitions that are being included in that.

DR. BERNS: Can I raise one other exclusion issue? That is that there is no upper-age limit to this, which, again, maybe there ought to be, in that an 85-year-old, a 95-year-old -- you know, I don't know where the numbers should be drawn -- this may not be an appropriate or necessary component of care. There is a lot of people with CKD in that age group.

DR. NALLY: I would be curious what the measure stewards say to that. I know, for instance, that the NHANES data, when they are talking GFRs, the cutoff is 85. Because above

that, I believe there are so few people that they potentially could be identifiable based upon other demographic information. So, I think they have an arbitrary cutoff of 85.

We have 1500 people over the age of 90 with CKD in the registry.

MR. JONES: We could take that up with the Work Group. It is a good thought.

CO-CHAIR CROOKS: I think we are getting off the topic of validity now and talking about specifications. I think maybe when you are 95, you seek --

DR. NALLY: Part of the exclusions --

CO-CHAIR CROOKS: Exclusions?

DR. PACE: Yes, I mean it is part of, do you have the measure that is going to really be appropriate in terms of identifying differences in quality?

CO-CHAIR CROOKS: Okay.

DR. PACE: And so, I think what is being suggested is, why include those patients

because maybe it is not quality care to do lipid monitoring over a certain age? I don't know if that is supported by the evidence.

So, the question is, what basis is there either to include or to exclude?

DR. KLIGER: I don't think there is any evidence about what age to choose, you know --

CO-CHAIR CROOKS: Right, a lot of questions, no answers.

DR. PACE: Okay. So, I think, is that going to be a point of contention, Joe?

DR. NALLY: No, not a point of contention. I am trying to bite my tongue here because we have a couple of papers coming out or being presented at the ASN.

But in the very old, sometimes this as a risk factor tends to melt away because we interpret it as all these other competing risks and the fact that you have to die of something. So, let's just leave it at that.

DR. PACE: Okay. All right. Well,

are you ready to vote on validity?

Okay, Tenee.

CO-CHAIR CROOKS: Okay. High, moderate, low --

DR. PACE: High, moderate, low, insufficient.

MS. RICHIE: And Lorien, validity?

DR. DALRYMPLE: Validity, moderate.

MS. RICHIE: Moderate?

DR. DALRYMPLE: Yes.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, that is 21.

We have 1, high; 13, moderate; 6, low; 1, insufficient.

So, it passes reliability and validity. I think we can go on to usability.

DR. PACE: Yes. Right. Okay.

CO-CHAIR CROOKS: Joe?

DR. NALLY: This issue is usability, correct?

CO-CHAIR CROOKS: Right.

DR. NALLY: And again, we have some information from the measure steward about the public reporting and uses in CQI. That's about all I can say.

The evidence is not greatly detailed. The various websites are referenced. And I must admit, I didn't go and check those websites.

DR. PACE: And I think, generally, all of these PQRS/PQRI measures, physicians are reporting; performance data are not publicly available. You know, if you wanted to go look up a physician's performance, that is not accessible at this time.

MR. JONES: The intent, though, is that they would be used on things like physician compare and meaningful use and things of that sort. So, they are geared up to be done, used for public reporting.

DR. PACE: Okay. Are we ready to vote on usability?

CO-CHAIR CROOKS: Hearing no

objections, let's do it.

MS. RICHIE: Lorien, usability? Lorien?

DR. KASKEL: I'm sorry, I said moderate.

MS. RICHIE: Okay. Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, I think that's it.

So, 2, high; 16, moderate; 2, low.

Okay, the next one is --

DR. PACE: And then, feasibility.

CO-CHAIR CROOKS: Yes.

Joe, did you have any comments on feasibility?

DR. NALLY: The statement is made that the data comes out of the EHR, which would be easy enough to check the labs.

My question for the measure steward is, how does one go about checking the documentation in the medical record that the results have been noted? As opposed to the

laboratory surveillance, which is reasonably easy by comparison, it may prove difficult to find at what point during that year the physician wrote a note that said, "Cholesterol checked and not controlled."

CO-CHAIR CROOKS: So, this is a question to the measure steward.

MS. CHRISTENSEN: I just want to clarify that the measure is actually patients who had it performed. So, right now, with the specifications there is no need to be able to tell that the doctor looked at it because most systems just don't have that capability right now. But it is something that we have definitely discussed at the measure development strategy level.

DR. NALLY: My microphone now works.

(Laughter.)

But the end of the descriptor said the CKD patients who had "a fasting lipid profile performed and results documented at

least once within a 12-month period". That's the fly in the ointment.

If that, indeed, exists, is it negotiable to come out? Or does the steward feel like that is a key component of the measure?

MS. CHRISTENSEN: I'm sorry, can you ask it -- are you clarifying whether it is the order or the result that we are looking for?

DR. NALLY: Well, the measure says you are looking for both, a cholesterol and a documentation that the cholesterol result was reviewed.

MS. CHRISTENSEN: I'm sorry, do you have a specification --

MS. AST: No, I think what we meant was, actually, I think they mean the same thing. Performed and documented just means it was performed and it is in the chart. I don't think it means reviewed, but it means documented, meaning it is in the chart.

Currently, it is not specified that the review has to be done. So, if that wording is confusing, we are definitely open to changing it.

DR. NALLY: You might think of another word other than "documentation".

MS. AST: That has been discussed, also, in our meetings, that that word is confusing, and we are in the process of removing it for many of our measures.

DR. BERNS: If I can ask maybe a related question of you, if a primary care physician or an endocrinologist, a diabetologist, or a cardiologist orders a lipid profile, and I don't but I could still be aware of it, how would that be sort of scored in this measure? Or, as is often the case, it may be an outside lab, not our lab at the hospital that does it, and I have a PDF floating around somewhere in our electronic medical record.

MS. CHRISTENSEN: That is an

excellent question. So, as long as you are aware that the patient had the test performed, then that patient meets the measure for you.

DR. BERNS: But you just said that there is no way to document my awareness of the laboratory.

MS. CHRISTENSEN: Yes, but if you are going to report on this measure, then you have to know whether you are aware of it or not. Does that make sense?

DR. BERNS: So, I would have to go through all of my charts to see whether I documented my awareness of somebody else's having obtained a lab?

MS. CHRISTENSEN: So, in a paper world, that would mean manual abstraction, yes. So, someone would have to find that result somewhere in the patient's chart.

In an EHR world, if you had either an interface into your EHR of lab results from somewhere else or a shared system between, say, different specialty offices and your

primary care office, or whoever you are as the doctor, as long as that result is accessible in your EHR, you're good to go.

It gets more complicated if you start having access to outside systems where you would need to go look at that. The integration is not necessarily there.

Does that answer the question? I mean, in the ideal world, the EHR, everyone would have an EHR and you would be able to see that Dr. Smith looked at this and acted on it.

DR. BERNS: Oh, we understand.

(Laughter.)

MS. CHRISTENSEN: Yes. Thank you. Okay. So, it is just not there yet, but, hopefully, someday.

DR. NALLY: I believe that concludes those remarks.

(Laughter.)

CO-CHAIR CROOKS: Did you learn something there you could share with us about feasibility? It is not as feasible as we

would like it to be.

DR. NALLY: We are moving forward.

CO-CHAIR CROOKS: Okay. Other issues related to feasibility? Questions? Concerns?

(No response.)

The reviewers sort of had mixed opinions about the feasibility, as I guess the conversation demonstrated.

DR. PACE: Okay. Ready?

CO-CHAIR CROOKS: I guess we are going to take a stab at it.

So, feasibility, high, moderate, low, or insufficient.

MS. RICHIE: Lorien, feasibility?

DR. DALRYMPLE: Feasibility, moderate.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Results: 12, moderate; 8, low; 1, insufficient.

So, I think it does carry it,

moderate.

DR. PACE: And overall.

CO-CHAIR CROOKS: So, overall, did we pass it? I can't remember.

DR. PACE: No. It's next.

CO-CHAIR CROOKS: I know. I know, but, of the four --

DR. PACE: Yes.

CO-CHAIR CROOKS: -- we passed all of them really.

Okay. So, let's vote overall. Yes, no, or abstain.

MS. RICHIE: And Lorien, overall?

DR. DALRYMPLE: Yes.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, 21 responses.

And it looks like the ayes have it with 18; 3, no.

And I think we have earned a bio break at this point, haven't we?

DR. PACE: Yes. Yes, definitely.

CO-CHAIR CROOKS: Thank you. Thank you.

Okay. Back in 10 minutes or so?

DR. PACE: Yes.

(Whereupon, the above-entitled matter went off the record at 4:14 p.m. and resumed at 4:29 p.m.)

CO-CHAIR CROOKS: We had hoped to end at 5:00, but we would like to ask to extend your time a bit. No, we are not going to be as drastic as Alan would if he were sitting here because he wouldn't let you off until we finished the agenda.

(Laughter.)

Well, how about 5:45? Is that reasonable? Forty-five extra minutes. I think we can get through a few more of these. Okay. All right. Well, thank you very much.

Okay. The next metric, then, is --

DR. PACE: 1633.

CO-CHAIR CROOKS: You've got my master list there.

DR. PACE: Yes, I'm sorry.

CO-CHAIR CROOKS: 1633, blood pressure management.

Dr. Fenves is going to take us off on this one.

DR. FENVES: I hesitate a little bit after this. I am new at this, and this is a thick one, but I will do my best. Just no laughing allowed, I hope.

(Laughter.)

It's like when I play golf; no laughing allowed.

So, this is a measure on the blood pressure management. The steward is the American Medical Association.

And it looks at the percentage of adult patients age 18 years or older with a diagnosis of Stage 3, 4, or 5 CKD, but not receiving RRT, and albuminuria, with a blood pressure either less than 130 over 80 or greater than 130 over 80 with a documented plan of care.

Now, in terms of definitions, first, the albuminuria is defined as greater than 300 milligrams of albumin, not protein, albumin for 24 hours. And the documented plan of care in those whose pressures are greater than 130 over 80 is recheck blood pressure within 90 days, initiate or alter pharmacologic therapy for blood pressure control, initiate or alter non-pharmacologic therapies such as lifestyle changes. Documented review of patient's home blood pressure log which indicates that the blood pressure is or is not well-controlled.

So, again, in this case, I already defined the numerator, and the denominator would be all patients age 18 years or older with CKD 3, 4, or 5, not receiving any form of RRT, and albuminuria as I defined it.

DR. PACE: Okay. So, let's go and start with impact, 1a.

Andrew, I don't know if you have any comments about that. It looks like the

group's preliminary reviews were pretty much thinking it was moderate or high.

DR. FENVES: Right. Of course, the impact is high in the sense that, first of all, hypertension is extremely prevalent, but, in particular, I think the issue has to do with, as all of us clinicians know, especially in CKD patients, whether their CKD is due to hypertension or worsened hypertension or to have secondary hypertension, obviously, in either case, blood pressure control is so intensely important with respect to preventing progression towards worst-stage and/or end-stage renal disease. So, that certainly appears that the impact is significant in that respect.

DR. PACE: So, is there any discussion about impact? Or can we go ahead and vote on that aspect?

DR. NARVA: Are we talking about the impact of blood pressure control or the impact of this lower target blood pressure?

DR. PACE: We are talking about, in general, the impact of the topic of blood pressure control. And then, we will get into the evidence about the specific target. Does that make sense? Okay.

So, 1a, impact.

CO-CHAIR CROOKS: Just for clarity, though --

DR. PACE: Yes.

CO-CHAIR CROOKS: -- the numerator includes not only patients who have high blood pressure, who obviously need something done, but also patients with meeting the goal, but just have the albuminuria present. And that could be patients, for instance, who used to have much higher albuminuria and, with blood pressure control, now it is much lower. That may be the best you can do in certain patients.

So, I am not quite comfortable unless maybe someone can explain to me a little bit more about why the numerator is

written in that way. What is the expectation that you are going to do for patients based on that?

Alan?

DR. KLIGER: But it looks like it means recording the blood pressure, making sure that it is recorded, for people with normal blood pressure, and making sure there is a plan of action for people with high blood pressure.

DR. PACE: And I think we should address that when we talk about the measure specifications and how it relates to the evidence. Because I think it is a question of, if you have either/or, is it basically everyone is going to be 100 percent? So that definitely needs to come up under specifications, if that is okay.

CO-CHAIR CROOKS: To clarify what I think I hear you saying, too, it is that, okay, so you're under control; your blood pressure is under control. You have

proteinuria. So, the plan of care is rechecked in three months. Okay.

DR. PACE: All right, 1a, impact, high, moderate, low, insufficient.

MS. RICHIE: Lorien, are you still on the phone?

DR. DALRYMPLE: I am.

MS. RICHIE: Okay.

(Laughter.)

Thanks for hanging in there.

DR. DALRYMPLE: Did I not say that enthusiastically? I will rephrase it.

(Laughter.)

MS. RICHIE: 1a, impact, high, moderate, low, insufficient?

DR. DALRYMPLE: Impact, high.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay. We have 21.

Okay. Twenty high and 1 moderate. Very good.

DR. PACE: All right.

CO-CHAIR CROOKS: That's what I like to see.

Okay. Is there a performance gap? Andrew?

DR. FENVES: With respect to performance gap, on page 4, they are looking at some data, not that surprising, from 2008. If you look at optimal blood pressure, say 130 over 80, or say less than 130 over 80, a substantial percentage of patients, 43 percent I think, did not meet that. I think we know that in clinical practice; these are patients with advanced kidney disease, often on multiple medications. There are compliance issues and the like. So, that is not surprising. So, there certainly appears to be a gap.

And then, they talk a little bit about ethnicity as well, in particular, in the African-American population.

DR. KLIGER: So, again, just to

clarify, this was from CRIC. These are data from CRIC showing these numbers that you just showed.

DR. FENVES: Yes. Correct.

CO-CHAIR CROOKS: Okay. More discussion from other members on the reviewing team? Or anyone?

(No response.)

DR. PACE: Okay. Then, let's go ahead and vote on performance gap.

Go ahead.

High, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: High.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: All right. The voting was 14, high; 6, moderate.

Now this is not a health outcome, right?

DR. PACE: Right.

CO-CHAIR CROOKS: This is a

process --

DR. PACE: Intermediate clinical outcome.

CO-CHAIR CROOKS: So, we need to look at the body of evidence.

DR. PACE: All right. So, Andrew, do you want to talk about the quantity, quality, and consistency of the body of evidence?

DR. FENVES: Right. So, in the body-of-evidence section, to be honest, they do a lot of dancing around. But, ultimately, if I am correct -- and they talk about review of the literature by the KDOQI group, a large group, which will come up again -- but if I read this correctly, it is a number of indirect studies looking at blood pressure lowering. Although I think in many of these studies, the goals are variable. So, it is not always 130 over 80, or there are some certain limitations. So, the data is good, but I mean this expert Committee rated it as

strong.

Those are kind of my comments. Maybe others can comment.

DR. NARVA: I have a question. Are we talking about the quantity of studies that justified this lower target? Or are we talking about the quantity of studies that justified blood pressure control to a higher target?

DR. PACE: No, this evidence should be specifically to support the measure as it is states --

DR. NARVA: There are virtually, there are very, very few studies. There is very little evidence to support either improved cardiovascular outcomes or improved renal outcomes when you go to this lower threshold.

DR. KLIGER: Again, specifically in the population that we are talking about here, Andy, of course, is right, there are no studies at these numbers. There is credible

information for the general population at this lower level. But in our patients I think that the evidence is not there.

DR. FENVES: Which is why I emphasized, I mentioned the word "dancing", and basically extrapolating from other populations and the expert panel, and so forth.

DR. NARVA: The lower target in the general population, what are you thinking of?

DR. KLIGER: The most recent HRV-7 or whatever, when I reviewed those data, there are studies that are clearly suggestive that that target is inappropriate and the quantity and quality of those studies, it would be reasonable to discuss, but it is a moot point for our patients because there are none for our patient population.

DR. FENVES: The only comment I would make is they did include proteinuria, I think because they were worried about that. And we all know that, say, overt proteinuria,

the way they define it, would be a worst prognostic indicator. But, again, I agree the evidence would be indirect.

CO-CHAIR CROOKS: I saw the developers hopping up and down. Do you disagree with this conclusion?

(Laughter.)

MR. JONES: We never would hop up and down here.

Andy, are you talking about folks with proteinuria as illustrated by the --

DR. NARVA: I mean with or without proteinuria, the studies aren't there. I mean the KDOQI that you are talking about is seven years old, and it is cited as the reference. But, within that, there is not the studies cited there.

If you look at the KDIGO which just came out, they make a similar recommendation, but it is 2c or it is not a high grade. And making this a performance measure, a potential performance measure, I think requires very

high-grade evidence.

DR. BERNS: I would also suggest that JNC-8, I think, is going to be coming out sometime -- oh, JNC-8 should be coming out relatively soon, I think, I don't know exactly when, which I think, from what I understand, will have very different numbers than this. Or at least it may make sense to wait until we see what that expert panel decides.

CO-CHAIR CROOKS: Will that address our population particularly or as a subgroup --

DR. NARVA: Well, actually, that is where I was yesterday and this morning. And I can't -- I'd have to kill you if I told you.

CO-CHAIR CROOKS: He would have to kill us if he told us.

(Laughter.)

DR. NARVA: It is a different process than the previous JNC processes. And I think that it is going to be very strictly evidence-based.

CO-CHAIR CROOKS: All right. Well --

DR. NARVA: And, you know, I don't know how interested this group is in harmonization, but it is going to be very strictly evidence-based.

DR. KLIGER: So, let me see if I understand you.

(Laughter.)

CO-CHAIR CROOKS: What can we read between the lines? Okay.

DR. PACE: And when did you say that is coming out?

CO-CHAIR CROOKS: November is what --

DR. NARVA: I mean there are a lot of people very impatiently waiting. What I heard is that at the American Heart Association meeting they are going to -- which is in November. But it has been deferred a few times.

CO-CHAIR CROOKS: Okay. So, are we

ready to vote?

Okay. First, on the quantity of studies in the body of evidence, quantity, not quality. High, moderate, low, insufficient.

MS. RICHIE: Lorien, quantity?

DR. DALRYMPLE: High.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Anyone else? Okay.

A hard spread. If you moved a couple from 1 to 3, it would be like even.

Okay. We have 8 voting high; 4, moderate; 2, low; 4, insufficient.

MS. RICHIE: Six, insufficient.

CO-CHAIR CROOKS: I'm sorry. Six voted insufficient.

Eight high; 4, moderate; 2, low; 6, insufficient.

Thank you.

Next, on the quality of the body of evidence, high, moderate, low, insufficient evidence.

Go ahead.

MS. RICHIE: And Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: So, 7, moderate; 3, low; 10, insufficient.

And let's vote, also, on consistency. High, moderate, low, or insufficient evidence.

MS. RICHIE: Lorien?

DR. DALRYMPLE: For consistency, low.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: All right. Seven moderate; 6 low; 8, insufficient.

So, applying our grid, where do we get --

DR. PACE: So, basically, we have inconsistent because it is either low or insufficient on that rating; plus --

CO-CHAIR CROOKS: Plus low quality.

So, that would give us a no, I think, the bottom row.

DR. PACE: Right.

CO-CHAIR CROOKS: Okay.

DR. PACE: Okay.

CO-CHAIR CROOKS: So, do we need to vote on the next question, which is the total importance?

DR. PACE: No, we don't vote on that because it won't pass if it didn't pass the evidence.

CO-CHAIR CROOKS: Okay. So, shall we stop here?

DR. PACE: Yes.

CO-CHAIR CROOKS: Yes. Okay. Good. All right.

So, let's go on to angiotensin converting enzyme inhibitor, or ARB, therapy and --

DR. NALLY: Can I make a comment about the last one?

CO-CHAIR CROOKS: You may.

DR. NALLY: And it is probably a followup of Dr. Narva's comment and timing of this report. You may have noticed I was listed as one of the people in that Work Group from many years ago, and you learn things with time, including more evidence and better look at the evidence.

If there are two bodies that are going to be commenting upon this subject, hypertension and people's kidney disease, both of which are about to come out with recommendations in the next several months, I would somehow encourage a resubmission. Because this is a high-impact question, but in order to not just simply harmonize with other groups, but also let this group actually look at the true evidence that is presented by the measure steward, that hopefully will be marshaled in such a way so as to make for a measure that would, indeed, be as correct as possible in today's evidence, and hopefully harmonize with other national and

international groups, hopefully, the proverbial minutes could recognize that and encourage a resubmission. Because it is a truly very important issue.

CO-CHAIR CROOKS: Could it come in through a different set of metrics?

DR. PACE: Well, I know that we actually have a cardiovascular project that was going on prior to this project. And I know that that project had some blood pressure measures, and I wasn't intimately involved there. Maybe tomorrow we can get some information for you.

Because I know that they were looking at blood pressure levels in general for the more general population, and I know there was a discussion about the JNC-8 coming out and where that was going to land.

But I will see if I can get some information to present to you. I guess I am not sure when the next opportunity would be, but certainly we do have some ad hoc review

processes. It is just that it is hard to keep doing ad hoc reviews. So, we would just definitely have to take a look at that, where we could fit that in.

But I hear what you're saying. Obviously, it would seem out of synch to come out with a performance measure based on some level that is then not supported by the major group that is reviewing the evidence. So, we definitely would want things to be in synch there. Okay.

CO-CHAIR CROOKS: Okay, 1662.

DR. PACE: And that was Andrew's as well.

DR. FENVES: For 1662, the same steward, the AMA, looking at percentage of patients age 18 years or older with a diagnosis of CKD -- I take this to be all CKD Stages 1 through 5 -- but not receiving RRT, and albuminuria, defined just like in the previous one, greater than 300 milligrams of albumin for 24 hours, who were prescribed

either an ACE inhibitor or ARB therapy within a 12-month period.

Again, that would be the numerator. The denominator would be basically everybody age 18 years or older who have CKD, not receiving RRT, and have albuminuria. So, again, I guess looking at the percentage of patients receiving either an ARB or an ACE, defined in that category. All-comers, not just diabetics.

DR. PACE: Okay. Any comments about impact? Any issues about that? Or can we vote?

(No response.)

All right. Why don't we go ahead and vote on impact? Then, we will get on to opportunity for improvement.

Okay, 1a, impact, high, moderate, low, insufficient.

Go ahead and start it.

MS. RICHIE: And Lorien, impact?

DR. DALRYMPLE: Impact, moderate.

MS. RICHIE: Thank you.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: A combo slide.

Okay. The voting was 13, high; 7, moderate.

DR. PACE: Even the program is getting tired.

(Laughter.)

CO-CHAIR CROOKS: Okay. Is there data supporting that there is a performance gap?

Andrew?

DR. FENVES: On page 3, there appears to be a gap. Let's see, I lost it now. But they quote data from 2008. But it is on dialysis patients and actually doesn't include this group because that's been excluded here.

DR. PACE: So, is the Committee aware of other information about performance gap for the broader population of this measure?

DR. BERNS: You know, again, we have an issue of heterogeneity here. This is virtually everybody in the world who has 300 milligrams of albumin or more for 24 hours regardless of their age, race, serum creatinine level, life expectancy to some extent. So, it is a very broad group, and their performance measure data only vaguely relates to ESRD, as far as I can tell.

DR. KLIGER: Right. Well, again, just to this specific question of a performance gap to non-dialysis CKD patients, the data that is cited is the USRDS dialysis population. So, there is no evidence we have been presented on a performance gap, none.

DR. DALRYMPLE: Can we clarify with the stewards about the statement there is a gap in care shown by the 2008 data of 44.9 percent of patients did not receive the optimal care, did not receive an ACE or ARB and had albuminuria?

CO-CHAIR CROOKS: Right. Can you

clarify? Is that in dialysis patients or is that in the target population?

Say that into a microphone, please.

MS. CHRISTENSEN: That data is presented for this specific measure. It is used in PQRI.

DR. LATTS: Yes, I think it is just confusing the way it is laid out because the dialysis and transplant patients is right above, but then they do say "this measure".

MS. CHRISTENSEN: There should be a line in between those two lines, yes. They are separate.

DR. KLIGER: So, what is the data source for the non-dialysis CKD performance gap?

DR. LATTS: PQRI claims. Oh, the 10th percentile is 11 percent, and the 90th percentile is 100 percent.

DR. BERNS: So, that data is three years old, I guess. And I would want to make sure that we could span the entire spectrum of

CKD at that albumin level.

CO-CHAIR CROOKS: What's that, Jeff?

DR. BERNS: The data are three years old. And I would want to make sure that it applies to this entire patient population, not only a segment of the patient population; that is, everybody with albumin level of 300 and above. And then, you get into issues of whether it is appropriate to have all of those patients on ACE or an ARB.

DR. LATTS: Well, I think it was this actual measure. So, it is the entire -- again, we can argue whether or not that is the appropriate thing, but it was the entire spectrum of CKD patients in this data.

MR. JONES: Yes, there are two different references.

CO-CHAIR CROOKS: Can you clarify this for us some more?

MR. JONES: Yes, there are two different references. One is the CKD data

from USRDS that I can't tell you whether it was broken down by proteinuria, but it was the CKD population there.

And the second one is the PQRS data itself is what is being referred to here. And that is in that table that is there with the breakout. You can see 44.9 percent of reported did not receive the optimal care. That is in those folks reported from PQRS.

CO-CHAIR CROOKS: Okay.

MR. JONES: But I can't tell you about the USRDS database.

CO-CHAIR CROOKS: So, for those of us who don't know these various databases and their nicknames, can we narrow it; have we been assured that this 44.9 percent of patients reported on who did not receive the optimal care, that is applying this metric to the target population, as described in the denominator? Is that true now?

MR. JONES: Yes.

CO-CHAIR CROOKS: Okay.

MR. JONES: But the USRDS database, I can't tell you whether those are people that were only proteinuric at that level.

CO-CHAIR CROOKS: Right. So, USRDS doesn't mean ESRD in this case? This is --

MR. JONES: It is CKD and ESRD is published there.

CO-CHAIR CROOKS: Okay. So, that issue has been clarified. Okay.

DR. NARVA: I think, doesn't the data in the USRDS describe people who are hypertensive with diabetes and CKD?

MS. AST: Not exclusively, from what I understand. The research that I did, there is all different kinds of patients. It is chronic kidney disease. With chronic kidney disease, there's all kinds of stats, and then, with ESRD.

CO-CHAIR CROOKS: Are you okay?

(Laughter.)

DR. NARVA: I think so, yes.

CO-CHAIR CROOKS: All right. So,

are we ready to vote on the performance gap? High, moderate, low, insufficient. Ready? Okay.

MS. RICHIE: And Lorien, performance gap?

DR. DALRYMPLE: Performance, moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Is 19 the appropriate number now?

DR. PACE: It looks like everybody is done.

CO-CHAIR CROOKS: Okay. All right. Another blend.

So, we have, under high, 2 votes; moderate, 14 -- is that what you see? -- okay, and low, 2; insufficient, 2.

Okay. So, let's go to the quality of the evidence or the body of evidence.

DR. PACE: Right.

DR. FENVES: So, the quality, again, to me, the key issue here had to do

with the problem that ACEs and ARBs are used in different studies. So, as we know, first of all, some of these patients don't have diabetes. The data is very good in diabetic patients, the use of ACEs and ARBs. I think that is fully accepted, although, as a reminder, there is better evidence for ARBs in Type 2, better evidence for ACEs in Type 1 diabetes.

Now the issue is to put this in the context of just regular CKD without diabetes and proteinuria. And there, the data are smaller studies, but, again, they go back to that Work Group that I talked about earlier, the same group from many years ago, KDOQI, looking at a large number of smaller studies showing improvement, again, in proteinuric CKD patients with respect to at least progression. So, that group rated the evidence as strong.

Those were my comments.

DR. DALRYMPLE: This is Lorien.

One part that was difficult for me

in regards to the evidence based on what was presented is at least most of the studies I am aware of in diabetes, there are a few normotensive patients; the vast majority are hypertensive patients. And it wasn't clear, based on what was presented, if these studies actually require hypertension in addition to albuminuria to receive treatment with an ACE, since the measure does not require hypertension.

CO-CHAIR CROOKS: Well, that's right, this metric doesn't call for hypertension in the --

DR. FENVES: Right. That's true.

If I may make a comment about that, as somebody having been involved in a study where we are looking at patients with significant proteinuria and some degree of CKD who have no hypertension, those are very hard to find.

CO-CHAIR CROOKS: I'll bet.

DR. FENVES: I mean that is not to

say that it doesn't exist, but that is a low number.

DR. NALLY: In Ed Lewis' captopril trial in Type 1 diabetics, there were 409 patients, and about 96 of those did not have hypertension. So, roughly a quarter. And they had quantitatively/qualitatively the same outcome, the same renal protection as those people who had hypertension.

And if you add kidney disease, in essence, the higher the creatinine, the more bang for your buck you got with an ACE inhibitor.

But hypertension did not discriminate, the presence or absence did not discriminate results. So, I don't know IDNT data that well along that question.

DR. DALRYMPLE: There have been studies looking at the normotensive diabetic, but I just wanted to make that comment. It is sometimes hard to comment on the quality of studies that are directly applicable to this

measure, as the vast majority out there I think include a predominantly hypertensive population.

DR. NALLY: Right. A lot of the diabetic microalbuminuria stuff, hypertension was not a requirement.

DR. NARVA: My understanding, and I am not sure if this is what Lorien is saying, the evidence is pretty strong for normotensive diabetics with more than 300 milligrams of albuminuria. It is not so good for people, diabetics with less than that. But it is very good for diabetics who are hypertensive or diabetics who are normotensive but have more than 300 milligrams.

There is not a huge amount of data on people who are normotensive and non-diabetic who have 300 milligrams of albuminuria, although I am sure most of us would look for some reason to put somebody on a RAS inhibitor, but --

DR. NALLY: You would have to go

back and look at the study demographics of the REIN trial and the OPRI trial, but I think you're right. I think most of those people have high blood pressure that got into those trials.

DR. BERNS: My recollection, also, is that most of the benefit has not been convincingly demonstrated below about 500 of proteinuria. I mean most of that is albuminuria, but the 300, I am not sure exactly where that number, what the data are to support 300 as opposed to some different number.

DR. NALLY: Yes, I don't know.

And then, the other question that is a corollary to that, they say 300 milligrams proteinuria daily, which infers you have collected a 24-hour urine and have that number. Was there any potential surrogate markers like a protein/creatinine ratio or albumin/creatinine ratio? Or does it require a 24-hour urine? Because if the requirement

is 24-hour urinary albumin excretion, you are going to have a surprisingly small number of people that get that --

DR. DALRYMPLE: It is a (telephonic interference) then because at least (telephonic interference) provided to us. There's a lot of different ways for trying to ascertain albumin in the urine, and, actually, diagnoses that aren't very specific are one of the ways to get counted in the denominator. There are some, I think, clarifications needed on this specification measure.

DR. PACE: Can you clarify?

DR. DALRYMPLE: Oh, do you want to do that now?

DR. PACE: No. I am going to ask the measure developer to clarify.

MS. AST: This is another case where we have continued working on the measures through public comment. And in later versions of our measures, we have a different definition for proteinuria that we would be

happy to apply to this measure as well. And we just have it as proteinuria defined as more than 300 milligrams of albumin in the urine over 24 hours or, No. 2, ACR more than 300 -- I can't read those. Can you read those for me?

MR. JONES: Micrograms for every milligram of creatinine.

MS. AST: Or, three, protein and creatinine ratio more than .3.

CO-CHAIR CROOKS: Okay. That sounds like a good improvement.

Okay. So, are we ready to vote on the quantity of studies in the body of evidence? Okay. High, moderate, low, or insufficient.

MS. RICHIE: And Lorien, quantity?

DR. DALRYMPLE: Quantity, I have moderate.

MS. RICHIE: Okay.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: All right. We

have 1 voting high; it looks like 19 moderate, and 1 insufficient.

Okay. Now quality. Are we ready? Okay. The quality of the body of evidence.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Twenty-one. Okay.

It looks like 1, high; 8, moderate; 1, low; 1, insufficient.

And consistency.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Okay, the same as last. One high; 18, moderate; 1, low; 1, insufficient.

DR. PACE: No, that says 17.

CO-CHAIR CROOKS: That's a 17?

DR. PACE: Yes, 17 moderate.

CO-CHAIR CROOKS: One, 17, 1, and

1.

So, I think we're okay here.

DR. PACE: Right.

CO-CHAIR CROOKS: We have a moderate or high for each of the three.

DR. PACE: Uh-hum, and we're okay on importance, I mean impact --

CO-CHAIR CROOKS: And gap.

DR. PACE: -- and gap. So, we can move on to reliability.

CO-CHAIR CROOKS: Okay. Andrew?

DR. FENVES: On the reliability, I think I am supposed to comment on the group of patients, the data sample, which I wonder if it is the same as before. It is looking at the data that they looked -- am I right? It's the same.

DR. PACE: So, the preliminary reviewers had mixed reviews about reliability. So, a couple moderate to high and a couple low. So, maybe we need to hear about the concerns on the low ratings about reliability.

DR. DALRYMPLE: This is Lorien.

I was one of the lows. So, I can say as to why I rated it as low. It was largely, in part, due to the e-specification, similar issues to the past where in the attached appendix a lot of the laboratory tests are not urinary albumin studies. They are things like pre-albumin, calcium (telephonic interference) albumin.

This is all on page 2, I believe. I don't know if you guys have that up.

But it is just to show that the denominator is currently including a lot of laboratory tests (telephonic interference) relevant or appropriate.

And then, I was also concerned about some of the diagnoses used for the denominator, including orthostatic proteinuria, lordotic proteinuria, and macroalbumin-positive, which is not the focus of the measure since it requires macroalbuminuria. As currently specified,

these patients would all be put into the denominator (telephonic interference) with those diagnoses.

Now you could argue accidentally pulling serum labs should be okay because you shouldn't have a value that is (telephonic interference) with the definition.

There was extensive (telephonic interference) of pregnancy codes. I think that was about 100 or so patients with pregnancy codes, some of which were postpartum conditions. So (telephonic interference) to determine how many of those would really reflect (telephonic interference) but I think that is a small number.

And then, last was something we had already discussed, which is what were we going to do about urine testing or time to collection as opposed to just 24-hour urine collection.

So, those were listings I was hoping to get from the (telephonic

interference) or the same approach we used in the past, which is divorced from the EHR specifications now.

DR. PACE: Okay. So, it sounds like you have identified some additional issues about those e-specifications. If it is the will of the group, we can proceed the way we did before; proceed with the measure, excluding the e-specifications, and ask those to come back to us with some clarification and crosswalk to make sure that they do, indeed, reflect the measure as it is specified, unless the developer has any clarifications they could provide right now.

Okay. Other comments about reliability?

CO-CHAIR CROOKS: This is the same sample used for other measures, I presume. Question for the developers: this is, again, interpatient reliability? In other words, more than one evaluator looking at the same chart and coming up with the same data? Okay.

DR. PACE: Okay.

DR. DALRYMPLE: This measure, like the others, will be implemented in different ways, depending on who is implementing it? So, some will be charts versus electronic versus CPT codes?

DR. PACE: That is what they are saying, yes.

DR. DALRYMPLE: Do we have any data on the reliability of the CPT II codes for this measure?

DR. PACE: I don't know if it was provided in 2a2; 2a2 we need to look at. I don't believe this has anything about CPT II codes.

MS. CHRISTENSEN: The measure hadn't been implemented in PQRI when we did the testing.

DR. PACE: Okay. But you're still specifying it with CPT II codes?

CO-CHAIR CROOKS: Yes, go ahead.

MS. ANDERSON: In terms of the

reliability, on the denominator exclusions it talks about the documentation of patient reasons for not prescribing ACEs or ARBs, such as patient declined and other patient reasons. In terms of specifications and the reliability of extracting that data, I have a question about that, and the reliability of them being able to extract it and remove it from the denominator. And maybe that is a question for the developers.

CO-CHAIR CROOKS: Do you want to ask the developers how they handle that?

MS. ANDERSON: How are you handling that, the exclusions from the denominators?

MS. CHRISTENSEN: That's a great question. I have a list of the verbatim documentation reasons for exclusion. Is that what you are interested in?

MS. ANDERSON: Just the reliability of being able to give reproducible exclusion data either through the electronic medical record or hand extraction, and if it is going

to be reliable.

CO-CHAIR CROOKS: I'll check here. For this measure, the exclusion rate that we found using the specified exclusions in the measure was 18 percent, and there were 13 discrete verbatim documentation reasons for exclusion which were reviewed by the Work Group, and none were found to be inappropriate reasons for exclusion.

CO-CHAIR CROOKS: Okay.

DR. NALLY: But that relies on somebody doing hard-copy review. Since one out of five people are being excluded for that reason, that may become an issue if you try to translate this into an EHR.

DR. DALRYMPLE: My understanding is there are CPT II code modifiers that you are going to use to identify exclusion as well. But those are really at the will of the physician, right? You are trusting that they are telling you they were excluded for the reasons you think, right? Do I understand

that correctly with the CPT II modifiers?

MS. CHRISTENSEN: Yes.

DR. PACE: Yes. The developer is saying yes. So, basically, the approach to the exceptions for this measure and the other measures is these broad categories that sometimes give examples, but are basically individual physician-defined in terms of whether the physician thinks they should exclude the patient based on a medical reason, a patient reason.

And so, it is a question in this particular measure we actually have some data. It is under validity, which is 2b2, or I mean 2b3.2 that I think Joe mentioned. So, that is on page 22 of the submission.

And they said the exception rate was 18 percent. So, that means 18 percent of the patients ended up being excluded for -- and then they gave some --

DR. NALLY: I appreciate and understand that. My point is that somebody

with a vested interest in doing this correctly reviewed the medical record with great intensity to come up with that. My personal medical record has failed to document my ACE cough for nine years now. And I can guarantee you my friendly bumpkin that hasn't put that in the chart hasn't developed a CPT code that I never heard of.

(Laughter.)

To say that we excluded that on the basis of joint stupidity, I mean --

(Laughter.)

My concern is, should this be implemented broadly, it is going to be difficult to document in the medical record what I consider to be a significant minority of exclusions in the 18 to 20 percent range.

DR. BERNS: Can I raise another concern with the exclusions? That is, as I read through these, if this is going to be done in a primary care doctor's office, an internist, or a family doctor, many of these

reasons for not putting patients on an ACE inhibitor or an ARB is exactly the reason they should be on an ACE inhibitor or an ARB. So, there is going to be a problem, I think, in making sure that this actually translate into quality.

MS. ANDERSON: I have another point for clarification. In the exclusion, they talk about patient decline. But the numerator is patients who were prescribed ACEs or ARBs. So, if a patient declines to take them, do you still include the fact that the physician prescribed it in the numerator and then exclude that patient out of the denominator because they refused to take it?

Clarification for that, please, maybe to the developers.

DR. PACE: I'm sorry, we can't hear you.

MR. JONES: We could get you how that was handled, but we don't have an answer at this moment.

DR. PACE: Okay. So, we have been kind of talking about both reliability and some validity issues, primarily around the exceptions with the validity issues. Should this measure go forward, it is also going to have to be harmonized with there are other NQF measures about ACE and ARBs. And many of those don't have those kind of open exceptions.

But I think it is important for you to weigh-in on that in terms of what impact that might have, you think that has, on the validity of this measure being able to accurately reflect quality of care consistently across all providers. So, I think that is part of the question that is on the table.

And then, the other question is, you know, if you accept that these can be individually identified by each physician, will it actually be available in the records that will eventually be used to extract these?

And maybe there are some other questions.

CO-CHAIR CROOKS: I think that is a good summary.

Did you have a --

MS. CHRISTENSEN: Yes. So, to answer the question about the exception for the patient refusing the prescription, the way the measure is calculated is, if the patient meets the numerator, an exception is not looked for. So, if the prescription was given to the patient and then the patient would meet the measure, they would have to have refused taking the prescription from the physician to be an exception.

Does that answer the question?

DR. PACE: Okay. So, we can vote, when you are ready, we will vote on reliability and validity. This might be an area where you could make recommendations about modifying the measure, if that turned out to be a concern in terms of the measure

going forward, if there are reliability or validity concerns about that.

Other discussion about exceptions or other issues about reliability and validity?

(No response.)

CO-CHAIR CROOKS: Okay. Then, let's vote.

DR. PACE: Okay. So, we will start with reliability, 2a.

MS. RICHIE: And Lorien, reliability?

DR. DALRYMPLE: Reliability, low.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Results are -- is that a 7, moderate?

DR. PACE: Yes.

CO-CHAIR CROOKS: Eleven 11; 3, insufficient evidence.

So, that one comes out low.

DR. PACE: Yes. Exactly. Okay.

CO-CHAIR CROOKS: Okay. Validity.

Further discussion on this? This was also face validity? Yes, we haven't talked about that.

DR. PACE: Yes, I think they did face validity. And then, the other aspect that affects this, as we talked about, are any concerns about the exclusions or exceptions.

CO-CHAIR CROOKS: Okay. Are we ready to vote on this?

Okay, 2b, validity, high, moderate, low, or insufficient evidence.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Low.

(Whereupon, a vote was taken.)

CO-CHAIR CROOKS: Again, we have 7 moderate, I believe; 12, low; 1, insufficient.

So, it may be, then, that we --

DR. PACE: That would not pass scientific acceptability.

CO-CHAIR CROOKS: Right.

DR. PACE: So, that would stop there.

CO-CHAIR CROOKS: It is a very complex metric; that's for sure.

DR. PACE: Right. It's already 5:30. Okay. So, it's already 5:30.

CO-CHAIR CROOKS: Yes, I don't know about you, but I am reluctant to jump into another one --

DR. PACE: Right.

CO-CHAIR CROOKS: -- and a whole new category, in addition.

DR. PACE: And we also need to have the measure developers to a brief introduction to those measures.

CO-CHAIR CROOKS: If we were to do that. And we still need to have another public comment period before we end.

DR. PACE: Yes. So, why don't we do the public comment? Maybe we will have, then, with the Committee just a brief debrief to get some ideas if there is something we could do to move faster tomorrow. Then, tomorrow we will continue on tomorrow.

We do have to start with the mortality measure in the morning because our statistical consultant is only going to be available from 8:15 to 9:00. So, we will have to do that at least first thing in the morning. Then, we will probably resume with the dialysis adequacy measures.

So, let's do public comment.

CO-CHAIR CROOKS: Okay. The floor is open for public comment and developer comment.

In the back?

MS. McGONIGAL: Hi. Again, Lisa McGonigal from KCP.

We again thank you for the opportunity to comment on the measures. We would like to use this afternoon period to comment on the mineral metabolism, patient education quality-of-life measures in advance of your discussion tomorrow.

So, for the mineral metabolism, KCP previously supported Measures 0255 and 0261,

which are measurement of serum phosphorous and calcium, respectively. But a review of evidence supplied by KCP members indicates that these measures are topped-out with performance rates on average of about 97 percent and up, regardless of dialysis organization type. As such, KCP recommends that these two measures be placed in NQF reserve status.

KCP continues to oppose Measures 0570, 0571, 0574, CKD monitoring of phosphorous, PTH, and calcium, respectively, because the measures are not harmonized with the corresponding PQRI measure and are less rigorous than the testing recommendations in the KDOQI bone and mineral guidelines.

KCP supports the following measures for public reporting only: Measure 1655, ESRD patients with PTH greater than 400 and not treated with a calcimimetic or vitamin D analog, and 1658, patients with PTH less than 130 and continued treatment with calcimimetic

or vitamin D analog.

With the patient education and quality-of-life measures, KCP continues to support the following measures for reporting purposes only: Measure 0320, patient education awareness, physician level; 0324, patient education awareness, facility level, and 0260, assessment of health-related quality-of-life in dialysis patients.

Thank you.

CO-CHAIR CROOKS: Okay. Other comments?

(No response.)

Okay. Thank you.

DR. PACE: If we could, if the Committee would just bear with us for just a few more minutes, we would like to just see if you can make any comments about our process or if you have any suggestions. We will be kind of thinking about what we can tweak a little bit for tomorrow.

I think part of this is just there

is a lot to these measures, and it takes a while to get through these. We will get as far as we can get and work with you through electronic communication and conference calls.

But if you would like to express any comments, frustrations, suggestions, we are all ears.

CO-CHAIR CROOKS: Alan first.

DR. KLIGER: Yes, this has only been slightly less painful than the dentist.

(Laughter.)

It really is wonderful because it really is having us all pay attention, and there have been wonderful comments.

So, my suggestion would be this: Karen, I spoke to you earlier today. Your comments in pre-examining these measures were frequently excellent, right on the point, making very pertinent concerns and observations about these.

What I wonder if it might not help us is in the technical aspects of these

measures, like validity, reliability, feasibility, if we could perhaps get a recommendation from you, based on your review of these criteria, as a starting point for our discussions. I think it might make the discussions more focused and perhaps a little more streamlined.

DR. PACE: I appreciate your comments that you found those helpful. It is not part of our process to have staff start with a recommendation. You know, we do try to do some preliminary review and identify issues and questions that we present both to you and back to the measure developers.

So, I think we wouldn't be able to do that tomorrow, but it is certainly something that I will discuss at higher levels at NQF. I appreciate that comment.

DR. KLIGER: Just a quick rejoinder to that, which is that there are clearly parts of this that you shouldn't be doing. There are parts that have to do with our assessment

of the importance, our assessments, you know, the final assessments, et cetera. But I am really suggesting picking out those technical pieces that we have spent a lot of time scratching our head about here, where we could get a head-start on that discussion with your observations.

DR. PACE: All right.

CO-CHAIR CROOKS: Jeffrey?

DR. BERNS: Two comments. One is related to Alan's. I wonder whether it would be worth looking at how the four or five people who did the measure review, if they, with some degree of unanimity, agreed that it didn't pass muster, could it just simply not come to this Committee? It might be worth going back and looking retrospectively at how successful that approach would have been. We may be able to eliminate going through all of these in some detail.

DR. PACE: Are you saying, based on the preliminary evals, if it didn't pass

muster --

DR. BERNS: Yes, so if three of the four said --

DR. PACE: Right.

DR. BERNS: -- maybe it doesn't need to come here.

CO-CHAIR CROOKS: You know, maybe they don't need to come to the full Committee.

DR. PACE: Right.

CO-CHAIR CROOKS: Or maybe the Chair could handle it or something.

(Laughter.)

DR. BERNS: Yes.

DR. PACE: Right. Well, I think that's one of the things that --

DR. BERNS: Or Joe.

DR. PACE: I appreciate that. One of the things we were talking about just with staff is that those preliminary reviews are valuable in terms of kind of identifying where the issues are. So, maybe that is one way to at least identify those that perhaps it looks

like they don't pass muster, but certainly open it up to make sure that there is agreement there versus that kind of moving too quickly.

DR. LATTS: Maybe, likewise, in that same vein, going through each of the various elements, if the preliminary reviewers have high agreement, everybody is high, we just ask if anybody has any issues and then go past it.

DR. PACE: Right. So, help me play that out. So, just use those ratings unless someone had a disagreement and not have to have the full Committee vote?

DR. LATTS: Where there is high agreement, maybe in the positive, maybe in the interest of giving the reviewers a fair review for the measure, if there was a negative, we go through it as a Committee. But where it is, yes, this was all valid, yes, this was our agreement, yes, it is consistent, we move past it.

CO-CHAIR CROOKS: And the presenter could be responsible for pointing that out more clearly.

I would favor maybe a format where -- at least I could put this out for consideration -- here is the metric. This is the presenter now. Here is the numerator, denominator, or in a sense what the metric was. In general, we found that this seemed to have high impact. They couldn't show us a performance gap, however, and we thought the body of evidence was okay. You know, and just kind tick, tick, tick kind of down and say this is where we all had agreement and this is where we had some issues.

It might keep things a little more together. It feels kind of disjointed. We just kind of jump right into one thing. So, that is a thought.

DR. NALLY: If you were going to go through that, my request would be to look at more advance prep time for that presentation.

Part of the concern here is stuff was coming at me while in the airplane last night. If the presentations have to be that representative and precise and fair, one should have adequate time for that preparation.

CO-CHAIR CROOKS: Yes, particularly knowing what your review mates thought, because you don't get that until very late.

DR. LATTS: Right, although that is not NQF's problem. That is us as reviewers.

DR. PACE: Well, just to clarify, what we sent you yesterday was the latest update, which was about seven more reviews. But, definitely, we understand.

So, I'm not saying -- you know, we will kind of regroup here and talk about it. We don't want to upset the process or not give things a fair hearing, given what we started with. But we just wanted to get some thoughts, and we will see if we can move things along tomorrow and we will get as far

as we can.

Yes, Janet?

DR. WELCH: I think part of this is just it is a new process. We have to know what the questions are. I think by having a group discussion, even though it is laborious, that that helps identify the questions.

DR. LATTS: I think Lorien deserves the real award for being on the phone all day.

DR. PACE: Right. Lorien, are you still with us?

DR. DALRYMPLE: I'm still here.

(Laughter.)

(Applause.)

DR. NALLY: Lorien, could you tell us exactly what type of Mojito you have been drinking?

(Laughter.)

DR. DALRYMPLE: That is the problem with a 12-day-old; there is no alcohol.

(Laughter.)

DR. PACE: Okay. Well, let's

adjourn for the evening.

CO-CHAIR CROOKS: Leave these on the desk, right?

MS. RICHIE: Please leave your voting remotes at your seat as well as any flash drives that we have given you today.

We have dinner reservations for you at 6:00 p.m. at M&S Grill. It's on the corner of 13th and G. So, one block up and one block over.

CO-CHAIR CROOKS: Which grill?

MS. RICHIE: M&S Grill. M&S Grill, on the corner of 13th and G. 6:15. I'm sorry.

So, if you walk out of the hotel, hang a left, one block up. We are on 12th and F.

DR. PACE: Are you going there?

MS. RICHIE: Yes. So, if you want to meet --

DR. PACE: So, do you want to meet in the lobby at six o'clock maybe?

MS. RICHIE: Uh-hum. Okay. You can head over now.

(Laughter.)

DR. PACE: She is saying that she needs a little more time. So, how about meet in the lobby about five after 6:00? And if you want to come on your own after that, you can ask at the concierge. They will know where the M&S Grill is. It is at the corner of 13th and F.

DR. FISCHER: What time are we meeting tomorrow? Is it still the same schedule, meet here at 7:30?

CO-CHAIR CROOKS: Yes. We will have to end on time because --

DR. FISCHER: No, that's fine. That's why I wanted to ask. So, are we meeting early?

DR. PACE: Let me ask, does anyone need to leave here before 3:15? Okay. So, we will end at 3:15. Continental breakfast will be here at 7:30, and we will start at eight

o'clock sharp.

(Whereupon, the above-entitled matter went off the record at 5:40 p.m.)