NATIONAL QUALITY FORUM

+ + + + +

RENAL ENDORSEMENT MAINTENANCE STEERING COMMITTEE

+ + + + +

WEDNESDAY
AUGUST 17, 2011

+ + + + +

The Steering Committee met at the Marriott Metro Center, 775 12th Street, N.W., Washington, D.C., at 8:00 a.m., Peter Crooks, Co-Chair, presiding.

PRESENT:

PETER CROOKS, MD, 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 & 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:
HEIDI BOSSLEY, MSN, MBA
TENEE DAVENPORT
KAREN PACE, PhD, RN
LAUREN RICHIE, MA

ALSO PRESENT:
KERI CHRISTENSEN, American Medical Association
EDWARD JONES, MD, Renal Physicians Association
DIEDRA JOSEPH, American Medical Association
LISA MCGONIGAL, Kidney Care Partners
JOSEPH MESSANA, MD, CMS
WILLIAM GOODMAN, MD, Amgen
XIA HE, Duke Clinical Research Institute
TIM KRESOWIK, MD, Society for Vascular Surgery*
ROBYN NISHIMI, PhD, KCP/KCQA
TOM NUSBICKEL, Amgen
JEFFREY PEARSON, CMS
ROBERT WOLFE, PhD, CMS
ELEFTHERIOS XENOS, MD, PhD, Society for Vascular Surgery*
IRINA YERMILOV, MD, IMS Health

*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

8:01 a.m.

CO-CHAIR CROOKS: Okay. So welcome back. Just to recap a little bit, yesterday we finished work on nine metrics, three of which were passed. That leaves only 25 to go. And I think it's obvious that we can't get through 25 and do a really good job in one day. Yet, that's all the time we have together.

So, Karen and I, and Helen and Karen and I have a new process in mind that I've agreed to. I think it will be better and she will explain it to us in a few minutes.

But before we go into that, I'd like to have Lauren kind of recap what happened with the last set of metrics that we passed at our last meeting in January. For those who were involved, maybe you'd like to know what's happened to our work.

MS. RICHIE: Good morning, everyone.

I know it's been some time since you last heard what happened with the last round of measures. So on July 13th our Consensus Standards Approval Committee, our CSAC as we call them, they approved all ten measures that were moved forward.

Now originally the Committee put forward 11 measures, but CMS since then withdrew their lower limit hemoglobin measure, so that made it 10. The CSAC approved all ten. The Board recently ratified the CSAC's decision, just last week it was. The press release has gone out. The measures are now endorsed. However, we have a 30 days appeals process for the measures, and that began on yesterday. So towards the middle of September we will have the appeals come in. We'll look at them again. Depending on what the appeals say and how many we get, we may have to go back to the Board and/or the CSAC depending on the content of the appeals. So after that we'll see what happens.

So, just to give you an idea.

CO-CHAIR CROOKS: So it can really take almost a year from the time we finish our work until the metrics have stepped through all the process and everybody's had a chance to give feedback and so on.

Kristine and I attended by phone the CSAC meeting, and it was interesting. While they eventually approved all of the ten metrics that were left, there was a lot of discussion. One on how distal the outcomes were to the outcomes we wanted, particularly the new pediatric metrics. And they were very concerned about that.

And what were some of the other big concerns?

MS. RICHIE: The frequency and assessment measures.

CO-CHAIR CROOKS: Yes.

MS. RICHIE: That was a major concern.

CO-CHAIR CROOKS: So as a heads up

to the Committee, they're looking for more and more proximal outcomes or the outcomes themselves.

DR. LATTS: Could I ask you a question on that?

CO-CHAIR CROOKS: Yes.

DR. LATTS: I mean we were too, and yet those measures are not submitted to us. So, you know obviously we didn't get what we wanted as a Committee. So how do we get what we want?

CO-CHAIR CROOKS: And that came up in discussion. They said well the Steering Committee doesn't write the metrics and we have to deal with what we have. And they did understand that pediatric nephrology had nothing and it's better to have something to start out then nothing. And they understand as a Committee we would have preferred to have been able to deliver better metrics.

Okay? So, I just thought you'd like to know what had happened and what will

happen with this work.

Okay. I'd like to ask Karen to describe a different approach to our work to try to make our time together as productive as it can be and yet give the metrics their full due.

DR. PACE: Okay. So I know this has been hard work for everyone, and we really appreciate you hanging with it. As Peter noted, we have 25 measures to go and, obviously, there's no way we're going to do that today continuing on in our process. So, I did confer with Helen Burstin last night, and certainly after your suggestions. And so what we thought could work is that rather then doing any voting today that we try to address each measure so that we can identify strengths and weaknesses, issues that need clarification, make sure that anything like that is fully discussed here. And then we will ask you to actually register your votes online after the meeting.

We'll give you the preliminary evals again and any of the discussion points from the meeting and then vote online. And then we'll have a conference call where we discuss the results of that voting.

The thinking is that, you know since we have you here collectively we want to take advantage of having you all here, things together, as well as we've got the measure developers here to do clarification. And so we thought that that would be the best use of our face-to-face time.

But I'll just stop there and see if anyone has any major concerns about that or if you think that would be workable?

Yes, Alan?

DR. KLIGER: I'm troubled by it. I'm troubled because the process that we've had has been one in which the voting is informed by the discussions that we've just had. And if we're going to have 25 measures or what fraction that are left that we're

going to be voting on remotely, touching and remembering and feeling the content of those discussions, I think will be difficult.

DR. PACE: Ruben?

DR. VELEZ: In that same direction I have concerns about doing it that way because the voting is the easiest. It's the discussion that takes time.

DR. PACE: Right. Right.

DR. VELEZ: So it's a lot easier if we have it fresh in our mind while we do this. That's my --

DR. PACE: And I hear what you're saying, but I don't see us being able to make things quick enough to get through even a substantial, and then we would have many, many phone calls to try to do that as well. So I hear what you're saying. I don't know.

Anyone else? Peter?

CO-CHAIR CROOKS: Well, the counterbalancing argument, though, is that we would have to go so fast and we would have to

be voting on metrics. And, frankly, speaking for myself I didn't absorb the full content of 34 metrics and their validity and all these arguments. And I think that it would be -- the product will be better because we will have given it a little more consideration and a little bit more time and not rush through it. So that's the opposite side.

I do recognize that it is a change in process and it is asking for, perhaps, a little more from all of you. But having committed so much to this process already, I hope that you'll be willing to do that.

DR. NALLY: Rick had an idea yesterday which in essence was a subcommittee phone call just before we come here. You already have us grouped by different --

DR. PACE: Right.

DR. NALLY: And what Ruben and I did yesterday at lunch was have a brief session of, you know this is yours; probably not good. This one, maybe this one's

discussible, et cetera. So there was a quick check where there was feelings of unanimity among the people that have reviewed them so that we could be on the same page. So that might really hasten the process.

DR. PACE: Right. And I --

DR. NALLY: And the other option I really think you have to consider if there has been so much energy expended on this, do we need to spend a third day here?

DR. PACE: Right. Would anyone spend a third day here?

So, I think that's an excellent suggestion and we can certainly try to work that -- you know have those subcommittee phone conference calls prior to the meeting. I think that's a good suggestion.

DR. BERNS: I would be inclined to agree with Alan and Ruben. I think we ought to do a really, really good job with as many metrics as we can and then leave the rest for another day rather then what I think would

force us to do a less good job with everything if we got through them the way that you suggest. And maybe we can figure out some other way to deal with whatever we can't get through today.

DR. PACE: Lisa?

DR. LATTS: Maybe -- I'm sort of of two minds on this. I don't know that we should just systematically go through these in order. I think we should prioritize either the easy ones and get them done or the controversial ones because I think those will benefit from a face-to-face discussion. And so maybe before we start for the day we should -- I know we have -- but maybe we should do a scan of the metrics that are left and try to prioritize.

But I do like Rick's idea not for us at this meeting, but for future meetings of having a subgroup meeting --

DR. PACE: Right.

DR. LATTS: -- ahead of time and

have them do that prioritization; yes that is clearly is out, yes this is clearly in, these are the ones that really need to be discussed in detail at the meeting.

MS. LeBEAU: I absolutely agree. Although, I think the easy ones are the ones that are easiest to do over the phone. Because for me it's great value being in the room with the more complicated ones that we really need to think through very clearly. So that would be my suggestion. I think prioritizing is a great idea.

DR. PACE: Okay. Well, why don't we take a poll?

DR. KLIGER: Can I just suggest that the easy ones are the ones that are reupping that have already been reviewed once before and for which there is just a -- you know, the additional amount to talk about what's happened since the last review. The harder ones are the ones that we're looking at for the first one.

DR. PACE: Okay. So let's get a pulse of the group and see whether you want to continue on as we did yesterday. So we'll put that forward or we can have the discussion, you know be sure that we address each of the measures today and then follow-up on line within a conference call.

So, I'll put forward the question of who is in favor of continuing the voting and --

CO-CHAIR CROOKS: So A is the original and B is the modified?

DR. PACE: Right. So we'll just do a show of hands since we didn't give you your remotes yet. But those who are in favor of continuing on as we were yesterday, raise your hand.

DR. DALRYMPLE: And Lorien is a yes.

DR. PACE: All right.

Is Max on the line? Max He? Okay.

CO-CHAIR CROOKS: He's due in a

couple of minutes.

DR. PACE: Okay. So I think what we will do then is we will start with the mortality measure. But maybe we'll just take a minute to identify some priority issues that we need to discuss.

So out of the measure, you know I think everyone would agree we need to discuss the mortality measure. It's complicated and there are some issues that we need to get resolved.

From the list of measures, are there any others that people would want to identify as high priority, you know based on your review?

There are also some issues with the older ones, though. Yes. But we could start with the news ones and then -- any other suggestion about new versus -- all right.

So is Max on the line?

DR. HE: Oh, yes. I'm here.

DR. PACE: Okay. Hi, Max. This is

Karen.

DR. HE: Karen.

DR. PACE: We're about to start. We're going to have CMS do a brief introduction of the mortality measure and then we will ask you to just maybe do a little presentation of the things that you provided in your statistical analysis. And then we'll have a discussion. Is that okay?

DR. HE: Yes, sure. Sounds good.

DR. PACE: Okay. So who from Arbor is going to -- okay. Bob?

DR. WOLFE: Bob Wolfe from Arbor Research.

And I understand that there are some issues related to the mortality that would be worthwhile discussing here. And I think it's a very interesting and important discussion which highlights the distinction between achieving the goals versus, maybe, following the standard practice.

So with regard to mortality,

mortality is a fundamental outcome so the questions of evidence and so on don't matter for mortality. But the real issue having to do with mortality is in the question of the adjustment for patient characteristics and the adequacy of that adjustment.

And, Lorien, if you could show the slide related to the different deciles. That's Figure 3. And this was sent to the Committee. And what it shows is how the mortality varies amongst the different groups of patients according to their predicted risk from the adjustment process.

Those of you who have the handouts, it is in Figure 3 from the analyses that we sent.

DR. PACE: It would be in the document that we sent the measure developer responses.

DR. WOLFE: Can you see it? That's it.

DR. PACE: And Max and Lorien, it's

in that measure developer response PDF. It's on page 33, Figure 3.

DR. DALRYMPLE: Thank you.

DR. WOLFE: What this shows is very widespread between the deciles of risk predicted and the actual mortality that is seen for those ten different groups from the lowest mortality with the highest survival at the top to the highest mortality or the lowest survival curve number 10 at the bottom.

I will say that adjustment for patient characteristics is always the glass half full, glass half empty. This is the good part of the story. There's a lot of discrimination between different patients with regard to their patient characteristics and our ability to predict the actual mortality that they will see. This is never a finished product in that we are always looking for new covariates, new factors that are predictive of mortality that can be and appropriately should be included in the model.

Some examples of that are given -- I'm not going to take you through it. But below there's some examples showing the careful modeling issues that have been dealt with with regard to BMI and also race by age, which we had in our model with an interaction for over a decade, similar to the Hopkins result that has just recently been published. But ours is not as pronounced and I am very interested in why it's a little bit different, even though it's effectively the same. But I think part of the explanation may come in what you'll see today.

The question before us is whether to adjust for race in this model. And let me explain why there are reasons not to. You may say well if it's predictive, you should always adjust for anything that's predictive. There may be reasons not to, and it has to do with a goal which was articulated by the NQF in a query to us, which is we do not want obscure disparities in access to quality care for

minorities.

So here's the problem: If minorities are getting worse outcomes for one reason or another and if we adjust for that, then we would say well that's just what's expected. So a facility that has lower -- worse outcomes for the minority patients would be okay because they would say well that's what we expect, that's what we see.

If you adjust for what you see, then that becomes the expectation and you say it's okay to be as expected. Are you with me?

So, facilities that treat a lot of minorities might have worse outcomes because they're giving, perhaps -- or minorities are getting poor care at those facilities at all facilities. But those facilities that have more minorities would have their outcomes excused because it's as expected. That's the problem, or at least as I understand it, that raises the concern about why we should or should not adjust.

If you adjust, you sweep it under the rug and say it's okay, it's as expected. That happens when outcomes for minorities are worse than for other patients.

What we have in ESRD is a different situation. In study after study, and this is not unique to our analyses, it has been seen that for whatever reason -- and I don't think anybody really knows the reasons, blacks on dialysis have better outcomes then whites of the same age.

The Hopkins results suggest that may be reversible or -- and but most blacks in the age range 40 to 70 and 80 have better outcomes. And it's substantially so. It's about 25 percent so.

So, I'd like you to move to Figure 1, if possible. It's just a couple of pages above there, Lauren. Thank you.

What this shows is mortality from two different models. And I want to focus first upon the red dashed line which shows an

unadjusted model where we do not adjust for race. And the mortality is shown on the vertical axis. And what we have done is grouped facilities into, I believe, ten different groups according to their case mix with regard to percent black.

The facilities on the right are those who have a high percentage black in their case mix. The facilities on the left are those facilities with a low percentage black in their case mix.

And what the red line shows is a general downward trend. It shows that facilities treating more blacks have better outcomes if you don't account for the fact that they're treating more blacks. They just do have lower mortality. My explanation for that is that's because blacks have lower mortality for whatever reason, and the facilities that have a lot of blacks consequently have low mortality because they have that case mix that does have lower

mortality. Just as facilities, if they were treating young patients, would have lower mortality then facilities treating old patients because old people have higher death rates than young people. Same here. Facilities that treat blacks have lower death rates because blacks have lower death rates.

Well, the question becomes then: Why is there that downward trend? I've given you one explanation. Another explanation is those facilities are better, and that's the naive interpretation that you would have if you just looked at that. Facilities treating more blacks have lower mortality, and maybe that's because they're giving better care.

In contrast if you adjust for race and say we expect better outcomes amongst blacks and then compare the observed mortality at these facilities to that expectation, then it turns out that those facilities which have low mortality because they're treating, I'll say patients who should have low mortality,

end up having higher than expected mortality; that's shown on the blue line. The blue line shows the adjusted mortality adjusted for race.

If you compare the mortality at those facilities to what would be expected given the fact that blacks are expected to have lower death rates, then they actually have higher death rates than you would expect for the mix of blacks that they have. And the facilities with few blacks have lower mortality than you would expect given their mix of patients.

I think it's really important to make sure you understand that. So, please, are there questions about those two curves? And it has to do with compared to what you would expect; either what you would expect given the race in blue or what you'd expect ignoring race in red.

DR. PACE: Before we jump in here, let me just ask -- the statistical review you

got from Max was before we got this response --

DR. WOLFE: I never saw the statistical review from Max.

DR. PACE: Pardon me?

DR. WOLFE: I never saw any statistical review from Max.

DR. PACE: No. I'm talking to the Steering Committee now.

DR. WOLFE: Oh, thank you. I'm sorry.

DR. PACE: Our statistical consultant.

So, Max, do you have any questions or any based on the response we got from CMS about the risk model or the race and ethnicity in the model?

DR. HE: Yes, I do have a question. So in Figure 1, the solid line, is that from the current model being submitted, the actual true modeling?

DR. WOLFE: The blue line is from

the model which is being submitted.

DR. HE: Okay.

DR. WOLFE: Which adjusts for the within facility race effect. We distinguish between between block and within block effects -- within facility and between facility effects and we are adjusting only for the within facility effect in the blue line. And that, we believe, clarifies rather then obscures the disparity in health care available to blacks because --

DR. HE: Yes. I totally agree. So minorities actually go to facility and they actually have better outcomes then adjusting for that and better differentiate between the facility. And in that case I'm looking at a perimeter coefficient from the Excel spreadsheet. And it seems that the blacks actually have worse outcomes, is that true?

DR. PACE: Right.

DR. HE: I'm looking at categorical black zero versus one.

DR. WOLFE: No. The reason it's complicated is because there are interactions of race with age, and that's been documented in quite a few studies. So it's important to put all of the factors involving race into the equation. There is no single number that compares blacks to whites in that spreadsheet that you have, but you have to calculate it for each age and then you'll see that actually blacks have better outcomes than whites at every age in that spreadsheet.

DR. HE: Okay.

DR. WOLFE: Okay. So that explains what appears to be this contradiction between these two curves. But it is because blacks actually have better outcomes on dialysis than whites, however that's not true for transplantation.

DR. PACE: Okay. We'll stop there for a minute and see what questions the Committee has.

DR. KLIGER: We always have to be

very weary of confounders when we look at data like this. And I wonder if you do a similar analysis for age, that is deciles of age and then units done exactly this way what that would look like?

DR. WOLFE: That's an excellent question. And the answer is if you had deciles of age here, the red line would go up and it does go up. That is facilities treating older patients have higher mortality because they have --

DR. KLIGER: Right. And then adjusted for age?

DR. WOLFE: Perfectly flat.

DR. KLIGER: Okay.

DR. WOLFE: Perfectly flat. Well, I'm sorry. It was closer to flat. It turns out that facilities treating older patients -- this is going to get complicated -- do better with older patients. Facilities treating younger patients do better with younger patients. So that actually the mortality came

down on both ends a little bit. And I'm not going to try and explain why that might be true, but it appears to be true.

I believe that -- go ahead, Jerry.

DR. JACKSON: This may be a naive question, but are there other risk adjustment formulas, models that would bring the blue line back to a ratio of closer to one?

DR. WOLFE: Yes. Another analysis which looks at the overall race effect including the effect of within facility and between facilities simultaneously attributes it all to race and adjusts for it and then it becomes flat.

The analysis that we have done tries to separate the facility effect, that is the between facility effects which is shown in the blue line from the race effect within facility so that you can understand what components of the higher and lower mortality are due to facility and which component might be due to race for whatever reason that is.

And I'm using race because it may be a socioeconomic effect, it could stand for lots of different things here.

I do think it's to go to the next figure, Figure 2, which is the same as the blue line in Figure 1 except it's broken out by race of the patients at each of these facilities. So again, the horizontal axis groups facilities according to the percent black. So facilities on the right are those in regions treating a high percentage of black patients, while those on the left are those in regions treating a low percentage of black patients. Actually, you'll see that ten percent of the facilities have zero black patients. There's a dot on the red line, an extra dot on the red line for those ten percent of facilities that have no black patients.

But in those facilities we then calculated the mortality for white patients, shown in red, and the mortality for black

patients, shown in blue. And what this shows is all patients fatality being treated at the facilities that treat a lot of blacks have higher than expected mortality compared to what would be expected for their race. And all patients treated at the facilities who treat a lot of whites have better than expected mortality for their race.

If you want to see disparities in health care, I think it's important to understand that this is what the adjusted analysis shows and what the unadjusted analysis shows. I will say, I am not trying to be a proponent of whether to adjust here or not, but I think that this Committee and I think CMS has to be aware of the consequences of adjusting or not adjusting in this rather unique situation where blacks have better mortality then whites.

I mean, we are the contractor to CMS. We are currently advice to CMS. We don't know what CMS will say about this

either. We just want to present the facts to you so that you can understand them and then make a knowledgeable decision.

DR. LATTS: Is this something that's known? I mean, is it known among the nephrology community that blacks have better outcomes then whites?

(Simultaneous speaking.)

DR. LATTS: Okay.

CO-CHAIR CROOKS: You might turn on your mic. But as long as my mic is on, I would say this is well known and in the research I've been involved with, which doesn't look at facility effect, but the age adjustment takes away the mortality advantage of blacks largely in other studies and not looking at facility effect at all. But it's pretty well known.

The prevalence of blacks on dialysis is about 3.2 times non-blacks.

DR. FENVES: I had one question, and maybe it's also naive, but when it comes

to transplantation for whatever reason one could make the argument that African-Americans are transplanted either at a lesser rate, at a different rate, they have immunologic issues. Now the question is if we adjust the transplantation rates, would this change? I mean, the point I'm trying to make is when you transplant the crème de la crème, the good patients and then unfortunately the patients who cannot be transplanted have a higher mortality for obvious reasons. So there's the question.

DR. WOLFE: So this is not a measure that's being put forward, but in fact the dialysis facility reports do report transplant rates. Those are not adjusted for race for exactly for the reason that you brought up. And this is an example where I believe that the solution that you propose might depend upon the particular situation that you're facing. And when there are disparities in a direction adverse to

minorities, you may make a different choice, perhaps.

DR. PACE: Lauren, could you bring up their spreadsheet with a coefficients or the comorbidity index?

DR. FISCHER: I have a question. This Figure 2, doesn't that seems to suggest that there's a strong facility effect independent of race? And I think this is very elegant the way this is done, and I think you nicely have laid out the argument that there it seems to suggest that their outcomes to some degree, how you look at the lines, are paralleling for why it's in African-Americans which there's something with the facility that is outside of someone's racial group which to me then would argue that probably adjusting for it makes --

DR. LATTS: But this should be published. I mean, if this is really not out there it needs to be published.

DR. WOLFE: The reason it's not

there is the separation of the race effect that would better race effect from the facility effect. And it wasn't until this question was raised that we actually looked at it in this particular way, although we had seen it before but had not published it.

DR. FISCHER: Because I think the question was this measure was supposed to be looking at a facility effect, right? I think therefore if you look at that curve, I think it shows that it's getting at the facility effect, which both races are paralleling with the facility effect. So to me then it seems like we should be adjusting for that. That the observed -- the expected formula is not unreasonable.

DR. PACE: Bob, could you just explain then on this table -- can you freeze the thing so we can see the heading? Is this the coefficients, the log of BMI? So I think in one of these blacks had a higher hazard then white. So I'm just trying to figure out

which table we should look at to see what the--

DR. WOLFE: So if those are the coefficients, and it looks like they are, there will be a coefficient for black. But since there are interactions with other factors, that will be the discrepancy for blacks versus whites for the reference group. And I cannot tell right now which is the reference group. And then that effect would be modified through its interaction with, I believe, it's both sex and age.

So the difference between black and white mortality depends upon the person's age and gender. So there is no single number that summarizes the enter comparison. And in fact, the way models are set up, the number for black will only compare for one particular subgroup.

I'm not sure if that addresses your question. And I'll let Jeff speak to this because he knows more of the details of this

spreadsheet.

MR. PEARSON: So I'll just note that the particular sheet you're looking at now are the mean values used for imputing the comorbidity index and the BMI. There's a sheet there on the bottom, there's I believe coefficients.

DR. PACE: Okay.

DR. WOLFE: Oh, so that was actually showing that blacks have higher comorbidity, is that right? Okay. Not that they have higher mortality?

DR. DALRYMPLE: Karen, can you clarify which spreadsheet we're looking at?

DR. PACE: It was in the folder with the information for measure 03669 and it was titled "SMR Models."

DR. DALRYMPLE: Thank you.

DR. PACE: That's the file. And we're in the worksheet labeled "Coefficients." Okay.

DR. WOLFE: And in this spreadsheet

if you look at line 18 "Race/Black," and that will be compared to the reference group of "White," the coefficient is minus .25. It is common to set up that coefficient so that that's a representative group. And I believe that that's what was done here. That's probably the typical age and it shows about 25 percent lower mortality for blacks then for whites at whichever age group this is. And we can look through this.

DR. HE: Sorry about this. The five column, is that zero versus 1 or what I'm finding under the "Black"?

DR. WOLFE: Yes. "Black" was coded as one for this particular covariant and the reference group "Whites" were coded as zero. The reference group was chosen as the largest group in order to give the most fatal estimates.

DR. HE: Yes. I read that.

So what is actually representing the categorical is that zero versus one so it

seems blind versus black, is that how --

DR. WOLFE: No, it's black versus white. Because there are separate dummies for three of the four different race groups. Black has its own indicator variable. Asian Pacific Islander has its own indicator. And Native American has its own indicator. So each can be compared to the reference group.

They can also be compared to each other by looking at differences between the estimates.

DR. HE: Yes. I don't understand part.

So are we looking at actually with the coefficients the and second column is high?

DR. WOLFE: Yes.

DR. HE: And there's categorical, so it says zero versus one. That's the only part that confuses me. So I think all you have been saying it should be one versus zero. You're comparing --.

DR. WOLFE: Thank you. I misunderstood your comment. And I thank you're correct. That would be more accurate and clearer. Yes. Thank you. That is black versus white.

DR. HE: In that case, when you present the black effect, what age do we use as the comparison group? Because I think there's a black age interaction, so you have to compare maybe three years of black and 40 years of white, is that right? What is the age point that you choose with this presentation?

DR. WOLFE: I would need to check to be confident. I believe the way the labels in the first column A are given that might be at age zero. But I'm not positive. There are age lines which are continuous linear functions. And I'm guessing that the reference group is set up as age zero. So that is not a very meaningful comparison. However, if you look at the lined plots and

figure, I believe it's five or six that we alluded to, they're relatively parallel for both blacks and whites.

DR. HE: Yes. Yes, I think if a patient younger then 18 years that are black has a higher risk, and for patients older then 18 years old patients has a lower risk. But I just want to make sure the direction to which the minorities are --.

And I think I totally agree with you if the minorities actually have better outcomes then adjusting for that will better differentiate between the facilities.

DR. PACE: Okay. Joe?

DR. NALLY: Bob, that's amazing data and I think I understand the questions and a profound observations have been made here. But I'm not a statistician that does spline plots and other things.

So, let me phrase the question this way: In my dialysis unit it's 91 percent African-American and my SMR is, say, 0.8

currently. And as I understand it the possibilities are either that's simply because I have a predominance of blacks or we could be providing better care, or both?

DR. WOLFE: That's if it were unadjusted.

DR. NALLY: So specifically that SMR right now is adjusted for race. And what you're proposing if it's not adjusted for race will it then answer the question better care or simply predominance of blacks? You know, how is the physician in the community going to interpret any changes we make here, and can that information be conveyed in an important way to address the primary issue of race and mortality?

DR. WOLFE: So right now the .8 is adjusted for race. So plausibly your mortality amongst your white patients is only 80 percent as high as for similar white patients across the country and the same for black patients. Actually, we don't know that

but that's the usual interpretation given to the .08 is it's .08 for all subgroups.

And the attribution, the appropriate interpretation is that's because you're giving good quality care.

If we had not adjusted for race, your SMR would probably be about .6 or .7 but we wouldn't know if that was because of good care or just because you're treating a lot of blacks. Either one could have lead to lower mortality.

DR. LATTS: The more relevant issue would be a facility that had an SMR of 1. -- it's those facilities that have a high percentage of blacks that would be performing well if it was not adjusted for race when adjusted for race, they would be performing more poorly and it's not reflected in the SMR because they're getting an advantage from having a higher population of African-Americans if it was not adjusted.

DR. FISCHER: Part of the question

eventually comes down to is if there is a survival advantage of African-Americans and there's an even distribution across facilities, how much of that is attributed to care or things being done at the facility versus something else unrelated to a facility effect? And I don't know if anyone knows how much of it's unrelated or related. A facility figure seems to suggest that there's a large component that is unrelated to facility effect. And if that's the case, then it seems more reasonable that that should be an adjusted part of the SMR.

CO-CHAIR CROOKS: Well, if a facility were to see both the race adjusted and the adjusted SMR, would that give them more information? Would that be clearer, more clear? That would help them figure out, you know is there improvement due to race mixture or facility effect?

DR. WOLFE: Rather then me answering that, let me ask you a reciprocal

question which may clarify it? Would it help you to see both an analysis which was adjusted for age and unadjusted? With the adjustment for age you would know that whatever excess or deficit mortality is compared to patients of similar age. Without it, you may see high morality and is that because you're treating old patients or because you have adverse care. You don't know. Without the adjustment, you can't parse it apart as easily.

DR. PACE: So you could give the results for a model with age and comorbidities without the face, or is that what you had already done?

DR. WOLFE: The red line is without adjustment for race in Figure 1.

DR. PACE: Right. But it did include age and comorbidities?

DR. WOLFE: Yes, it did. Thank you.

DR. VELEZ: I mean, this is amazing. When you look at data, in fact all data, it brings back some of the thought

process from 10, 20 years ago. And we realize how important some local factors, facility factors race, age, even transplant factors get involved and it's all very local.

Trying to get realistic in all of this, I have a worry in that this will require a collective thinking process change completely; networks, I mean, the whole nation. Because we've been using this rule. I mean, we've playing a sport and now we're suddenly saying okay, we're going to change the rules of the sport. And I wonder on reality check here is I think we need to move this forward. We need to start moving the process into changing our collective thought process, but I'm not sure we can do that here in the measures we're doing.

I mean, I'm now confused and concerned about how we may adapt this to what we're doing.

DR. LATTS: I actually don't think we should make any changes. I think we should

continue to produce SMR adjusted for age. I think if any change, we should give facilities that second table that shows them their adjusted mortality by race, which is potentially actionable as opposed to this which is not actionable, and I don't think very helpful.

DR. KLIGER: Yes, I agree.

I mean, Ruben, I don't think this is -- it's a great new view, but it doesn't change the way that we've been doing it. It endorses in my mind the strength of continuing to adjust for race in addition to age in comorbidities.

DR. FISCHER: If the logic has been that that there are differences in mortality by gender, race and age and while some of them may have to do with provisions of a care facility, a lot of them don't have anything to do with it. I think if we think about that in terms of age and gender and there's data about face, to then make an exception and to stop

adjusting for race, I don't understand why we would want to do that.

CO-CHAIR CROOKS: Okay. I think that closes that topic for me.

DR. PACE: Okay. So why don't we then we'll proceed through evaluating this measure. Who did we have assigned to present this measure?

CO-CHAIR CROOKS: Jeffrey Berns.

DR. PACE: Jeff Berns. And we can walk through.

Do you want to change your mind on the voting thing, Jeff?

So I think we can quickly go through the first ones here, unless you have something to say about impact. Shall we go?

Any comments before we just go to vote on impact? Okay.

Can I go ahead and start the clock?

CO-CHAIR CROOKS: High, moderate, low, insufficient.

MS. RICHIE: Lorien, impact?

DR. DALRYMPLE: High.

MS. RICHIE: Thank you.

DR. FISCHER: I'm actually presenting this?

DR. PACE: Oh, okay. I'm sorry.

DR. FISCHER: But wait, before I get up, but I'm happy to turn it over to my senior colleague.

CO-CHAIR CROOKS: Yes, let's keep it this way all day, right? Let's just roll along.

Twenty-one high, nobody moderate, low or insufficient. Okay.

DR. PACE: Okay. So now we will go to opportunity for improvement. And, Michael?

DR. FISCHER: And I think there was general consensus. I don't know if you can pull up the Excel spreadsheet, but among the five of us who reviewed this they had kind of presented that there was variation of facility by this measure. And that there was need for improvement overall. So I think all of us had

given 1B, it was, was a medium or a high.

The big issue which we've kind of been discussing for the last 15, 20 minutes was the issue about disparity data. And that went into this whole thing about adjusting that as to race. I won't rehash that. But putting that aside, everyone else thought that there was some variation by facility and therefore, opportunity for improvement.

DR. PACE: Comments from the other the other assigned reviewers or --

CO-CHAIR CROOKS: All right then let's vote on the performance gap. High, moderate, low and insufficient.

MS. RICHIE: Lorien, performance gap?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: Okay. Eighteen high, three moderate.

So this is a health outcome?

DR. PACE: Right. So on this one all we need to do is there plausible

relationships to health care processes and services that affect mortality?

DR. FISCHER: And they do that later in the application, Karen. They kind of point out -- I think it was in hemoglobin and anemia and Kt/V or URRs, from what I recall. But they had linked that with SMR.

DR. PACE: So -- yes?

DR. KLIGER: Can I just explore for a moment that there are those correlations. Is there any evidence that affecting any of those measures effects this outcome?

DR. FISCHER: Yes. I think that there were correlations given. I don't remember that they had actually formally looked at that if you made a modification in something as an intervention, that that changes SMR. I thought they were epidemiologic relationships but I can be corrected. But that was my recollection from what was put in the document.

DR. KLIGER: Can we ask the

developer that question?

DR. WOLFE: So, actually, we've looked at it the other direction. Maybe it's just what you're saying.

We have looked at specific practices and seen whether facilities that carry out one practice have different mortalities then facilities that carry out other practices. And the answer is very clear, and that's the strongest relationship that we feel we can document that's likely to get as close as possible to a randomized controlled trial is differences between facilities.

For example, that kind of analysis does replicate the randomized control clinical trial results for EPO showing that up to about 12 -- at above 12 you do get the higher mortality when you look at it in relationship to the standardized mortality. So that's a modifiable -- several modifier factors such as vascular access, adequacy of dose and anemia management all are related to mortality. And

I'll leave it to you to tell me which ones of those are modifiable.

DR. BERNS: I think Alan's question, if I'm understanding it correctly, is whether somebody has shown prospectively whether changing some pattern or practice changes SMR?

DR. WOLFE: And we have not replicated that with the Medicare data. All we have been able to do is look at practices that did change historically and correlate that with changes in outcomes. Other individual studies have been prospective in nature and have yielded similar results is my understanding.

DR. PACE: And we'll look at that more closely at validity in terms of can you make conclusions about quality based on that. At this level you can also look at the studies of treatments and treatment interventions at the patient level; does it effect mortality in terms of whether there are

health care practices that can influence patient survival or mortality rates.

CO-CHAIR CROOKS: Well, isn't it true that a given facility tends to do the same year after year; that a high performer tends to be a high performer and a low performer -- I think is sort is evidence, it may be indirect, but that there is a facility effect and that there is -- Alan's over there shaking his head no. I mean it's not the same as having a prospective clinical trial.

DR. KLIGER: I mean, at this level we're being asked whether there's a rationale that supports the relationship. And I personally from what I've heard think there surely is a rationale. I think that digging deeper into causality is something we need to do. But at this level, I'm comfortable with the relationship.

DR. FISCHER: It's been linked to intermediate outcome measure. Intermediate outcomes that are modifiable, right?

Hemoglobin and URR, K2/v. I mean, albeit that the strength out of the evidence is borne out of retrospective analyses of existing data.

DR. LATTS: And I guess my question is can a facility that a poor performance in SMR take action to improve it?

DR. KLIGER: That's the whole question we're asking here. And there is not a clear answer, although the data that they've analyzed would suggest that the possibility is yes.

CO-CHAIR CROOKS: So should we formally vote on this question?

DR. PACE: Right.

CO-CHAIR CROOKS: Okay. 1(c), health outcomes. So if the measure is a health outcome, does a rationale support relationship to at least one health care structure process, intervention or service? Yes or no.

MS. RICHIE: And Lorien? Yes or no for health outcome?

DR. DALRYMPLE: Yes.

CO-CHAIR CROOKS: Okay. Twenty-one, the magic number.

DR. PACE: Okay. So let's move on to --

CO-CHAIR CROOKS: That was 21 yes for the record.

DR. PACE: Okay. So let's talk about reliability and then we'll get into validity. So, Michael?

DR. FISCHER: So the reliability, they kind of talk about that they have standard sources for death, and then they also kind of described in terms of the expected, the Cox model which we've kind of talked about at length already this morning about what's included in the Cox model.

I think the one thing that was raised by myself and other people, and in the staff notes, was the idea that the reliability -- and we an ask the stewards for clarification, I think they may have

responded a little bit to this in one of the documents, is their initial approach to reliability was looking at SMR from year-to-year as a way of assessing reliability. And I think concerns were raised about is that really answering the question of reliability, that type of methodology in the measure.

DR. PACE: And, Lorien, if you could bring up -- right. They did some signal-to-noise analysis for the process measures but not this outcome measure. So maybe we could have the developer -- I don't think it was in there.

DR. WOLFE: No, we did not do the signal-to-noise racial analysis for that. But there are very substantial differences in the SMR from facility-to-facility. Typically within a random effects estimation of the variation, I got plus or minus 15 percent with regard to mortality. So, that's a substantial amount, a clinically important amount of variation that the measure identifies.

The motivation for putting in the serial correlation from year-to-year was we were thinking of that as a pseudo experiment of having two different raters rate the same facility. And all we can do is look at it in one time period compared to another time period, very close to it so they're independent evaluations but based upon different data. And the answer is that inter-rater reliability is quite high based on that correlation. That was the logic behind that motivation.

DR. FISCHER: I understand that. I guess the flip side is you believe what Alan said that if my facility got a bad SMR and hopefully I've done something, right? A process change -- I'm just trying to be devil's advocate. If I've then done some process change that hopefully impacts this outcome, that maybe my SMR would change a bit more from year-to-year over some time period, right? Depending on effective we are.

But I understand the idea that if we think that these things on the other hand are rather stable and that change is more insidious, then looking at inter-rater reliability from year-to-year is an unreasonable.

DR. PACE: Right. I think the concern of looking at that as reliability is that it's also different time periods and different patients even. And so even from that standpoint of trying to do it as a pseudo, it's really measuring something else.

CO-CHAIR CROOKS: Right. Well, the fact though that the data is managed electronically, you know at the level element, reliability it should be okay.

DR. PACE: Right. So at the data element reliability it's probably -- I mean --

DR. FISCHER: No. The data source is for death. I mean, the Master Death File --

DR. PACE: Right.

DR. WOLFE: -- and the Death Index I think are widely used data sources. They're imperfect, but I think they're fairly robust.

CO-CHAIR CROOKS: The death data, is it the forms that are filed or do you use these other ways to search for death? It's facility reported deaths, right?

DR. WOLFE: It's mostly reported deaths through the facility from the death forms reported by facilities, but it is supplemented by the Social Security Death Master File, which increases about -- that's where we also get about 10 or 15 percent.

As a final step, the data are put up for facility review before they are made public on the DFR. And actually, several facilities look at patient-by-patient lists of their patients to clarify and verify that the data are entered correctly. So it is actually done at the facility level in addition to what is originally submitted.

CO-CHAIR CROOKS: And that, of

course, is in their interest to do a good review. That's why I think that's another form of reliability check, isn't it?

DR. PACE: Right. And I guess the other question, since it's now so prominent in the risk model, is do you have any idea about the validity of the race data? And that's a validity question and I should probably hold that.

DR. WOLFE: Yes. It has been looked at and I don't know the right answer.

DR. PACE: Right.

DR. WOLFE: But here's what I do know. Is that there are standards for how race should be reported. It should be done as self-reported and there are certain categories that should be included in the race specification.

Right now the data are taken largely off of a 2728 form. And I believe that has recently been modified and, Jeff, you may be able to speak to this better than I

can. It's supposed to now reflect self-reported race, I believe, right?

MR. PEARSON: Yes, I don't think that has been implemented just yet.

We have done studies comparing different sources of race and ethnicity data that we have. So we compared to the UNOS transplant data and we've compared to the Medicare Enrollment Database. And we found very high agreement on ascertainment of white versus back. The other categories a little less so because it's provider report, but we have seen high agreement there.

DR. FISCHER: We looked at this in VA. I mean distinguishing between white and non-white is always pretty good with self-report. It's when you get to finer categories, Hispanic and Asian that there's more problem. But the white/non-white is usually pretty good.

I think the other thing about the 2728 data, right, is that the comorbidities

and some of the other data elements from it do suffer from under reporting and some problems. But that's a separate issue.

DR. MESSANA: Just one last bit of clarification of Jeff's comment. The comparison between white and non-white from the Enrollment Database and 2728 data sources is available in print in a American Journal of Kidney Disease article by Roach from 2010 which corroborates the high correlation between categories of black versus non-black. But that those reflect some of the greater difficulties in differentiating between other ethnic and racial groups.

CO-CHAIR CROOKS: Okay. Are we ready to vote on reliability?

DR. PACE: Any other comments from the other reviewers? Questions from the Committee? Okay.

CO-CHAIR CROOKS: Okay. So let's vote on reliability; high, moderate, low or insufficient evidence.

MS. RICHIE: And Lorien, reliability?

DR. DALRYMPLE: Moderate.

MS. RICHIE: Thank you.

CO-CHAIR CROOKS: Twenty-one. Okay. Seven rated it high, 14 moderate, none low, none insufficient.

So moving on to validity then.

DR. PACE: And this would encompass the validity testing and the risk adjustment model we've talked about. And, Michael?

DR. FISCHER: Yes. I mean, I think some of this we've kind of talked about, and there were some concerns. I mean, part of the concerns related around kind of the risk model testing and the modeling and the factors included in the models which we've kind of discussed at length.

You know, they related SMR to anemia and UR, these other measures, these well recognized intermediate outcome measures. And they showed kind of concurrence and

correlations which seem to indicate that SMR is robust. A lot of it I think hinged upon what we kind of discussed up to date, which was what are we all including in the models in the covariant section and how well is that giving us kind of what we assume is the expected outcome.

I think in general I was trying to look back at the spreadsheet. I think in terms of the voting, I think most of us -- I think most the people on here -- it's a little bit hard to see. Sorry, the spreadsheet's kind of wide.

DR. PACE: Yes. Actually, it looks like --

DR. FISCHER: I can't see it.

Okay. So it looks like everybody-- I think there was an insufficient. The rest were medium or high. I think the insufficient probably or a little bit individual. But I think that might have been related to some of the questions that we had had that we've kind

of discussed here to time.

CO-CHAIR CROOKS: I was a little bothered that validity was stated, too, because it showed some correlation with some other outcomes, and therefore it's a valid measure. I mean, how do you view that?

DR. PACE: Well, you know, for process measures that's great showing the correlation to outcomes. It's kind of, I guess, a question for all of you when you're looking at showing validity of the outcome measure what's an appropriate test.

DR. FISCHER: I mean, I think the two parts of this measure writer observed deaths and expected deaths. Observed deaths I think we probably agree that the sources being used are quite valid in determining observed deaths. I think expected deaths got to the whole discussion that we've already had about the model and what's included in the model. And essentially that is how are we coming up with a value for expected deaths. And I think

we've had a long discussion about that. You know, there are things that are just not known at this time. But I think that seems to be that you're looking at the face validity of the measure, and in this one the two parts are the observed and the expected deaths.

DR. PACE: So it seems like we've talked about some, like you said, the validity of the death data especially.

DR. DALRYMPLE: This is Lorien. Can I ask a minor question just for clarification? One of included adjustment variables is age adjusted population death by state and race. But it's based on the U.S. population in 2001 to 2003. Can you just clarify why that date is still being used and if that will be updated soon?

CO-CHAIR CROOKS: Did you understand the question.

MR. PEARSON: Yes. I believe that might be an outdated reference.

DR. DALRYMPLE: Okay.

DR. WOLFE: It is true, however, my understanding is that the data are lagged by more than a year or two because of reporting through our data source. However, the death rates by state and age are very stable over time, certainly over a few years period. We have worked as hard as we can to get the most current data available on that, but it is not as old as you've identified there.

MR. PEARSON: So the source for that is the National Center for Health Statistics a health publication that they put out annually that use the latest data released each year.

DR. DALRYMPLE: Okay. So it's probably not the 2005 data?

CO-CHAIR CROOKS: Okay. Other issues around validity before we vote? Okay. Then let's go ahead and vote. The usual scale, high, moderate, low or insufficient evidence.

MS. RICHIE: Lorien?

DR. DALRYMPLE: For validity high.

CO-CHAIR CROOKS: So five voted high, 16 voted moderate. So I think we can go on to useability.

DR. PACE: Yes. I think we don't need to talk about disparities in this one.

CO-CHAIR CROOKS: Yes.

DR. FISCHER: I think quick work of useability, this has been a previously endorsed measure. It's publicly reported. It's using dialysis reports. I don't think anybody, unless someone does now, I don't think any of us have concerns about it. So we can just move forward.

CO-CHAIR CROOKS: Well, from a QI front I'd say it's hard to know if you happen to have a low score, exactly what to do about it. But it is still I think a good process. So I think it's a little less useable for PUI then it is for public reporting, but it's still useable.

DR. LATTS: And actually my only

comment on public reporting is that I think a very large percentage of facilities are as expected with a relatively small above or below expected the way its listed. So it would be nice to have a little more differentiation from a consumer standpoint. I don't know if you guys looked at the stats.

DR. KLIGER: Yes. Only if that more differentiated was meaningful. So you have to be careful.

DR. LATTS: Right. Right. Yes. Agreed. Agreed.

CO-CHAIR CROOKS: So are we ready to vote for useability? Going to put both public and QI into one question, okay?

We'll vote high, moderate, low or insufficient. Go ahead.

MS. RICHIE: Lorien?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: And we have 15 voting high, six moderate.

So on to feasibility.

DR. FISCHER: I think similar to useability, overall feasibility I don't think is much of a concern. I think all of the reviewers, including myself, rated this as high. I don't know if any new concerns have come up, but that's what the preliminary evaluations were.

DR. PACE: Okay. Let's go ahead and vote on feasibility then.

CO-CHAIR CROOKS: Go ahead.

MS. RICHIE: Lorien?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: The votes were 20 high and 1 moderate. So overall, this measure meet all the NQF criteria to be suitable for endorsement.

Let's go ahead and vote. One yes, two no, three to abstain.

MS. RICHIE: Lorien, overall?

DR. DALRYMPLE: Yes.

CO-CHAIR CROOKS: We have 21 yes, zero no.

Thank you.

DR. PACE: Okay. What we're going to do using your suggestion about priorities with new measures and also some timing issues is we have some new measures under mineral metabolism and the developer is here this morning. So we'd like to at least have that advantage.

So what we will do is -- let's see, which ones are they. Is it 1655? We will go to 1655 and 1658, those are the Amgen measures on parathyroid hormone. And why don't we have the presenter. Would you introduce yourself and then just briefly give an introduction to your measures?

DR. GOODMAN: Sure. Thank you for the opportunity to speak this morning.

My name is Bill Goodman. I'm a clinical research medical director with Amgen.

We have put forth two measures with respect to PTH monitoring that we think are important from the perspective of patient

management and patient safety.

Several things that happened in the recent years that raised concern about the management of the secondary hyperparathyroidism in this population. This is a progressive disorder. Its severity increases over time. And it's been documented repeatedly in the literature that the severity of disease and ultimately the need for parathyroidectomy to manage it surgically is dependent on age, duration of chronic kidney disease or length of treatment on dialysis or dialysis vintage. So these are consistent predictors of the disease severity and its progression over time.

With the development the new KDIGO and KDOQI guidelines some additional uncertainty has been introduced. Secondary hyperparathyroidism is incorporated into this broader syndrome of chronic kidney disease, mineral and bone disorder. And the attention that the disease in secondary

hyperparathyroidism and its progression has somewhat been obscured.

Additionally, the KDIGO and KDOQI working groups set forth thresholds at the upper and lower end for PTH that they designated as depicting areas of extreme risk. Unfortunately, most of the broader community have interpreted those ranges as target therapeutic ranges in implementing updated practices guidelines.

So what we have suggested on the monitoring of disease progression relates to measurements of PTH that exceed a value of 400. In our submission whether one looks at the populations using large dialysis provider databases or DOPPS data, the percentage of patients with values above 400 ranges from 20 to 40 percent. And many of those individuals are untreated.

Additionally, if one looks at a facility level again a substantial proportion of patients approaching 40 percent have

elevations in PTH and nearly half are untreated.

So, it's our contention and recommendation, and we feel it's consistent with the KDIGO and KDOQI guidelines that recommend that PTH values be monitored and that trends, particularly upward trends for patients with values in the 300 to 600 range be identified and that the interventions to prevent those values from exceeding the upper threshold of 600 which defines a level of extreme risk in the KDIGO's view be considered.

On the lower end for PTH this represents a somewhat different population and many of these individuals do not have the disease of secondary hyperparathyroidism. Generally speaking these individuals are older, there's a high prevalence of diabetes, malnutrition is common and some of these individuals may have undergone parathyroidectomy in the past. So clearly they

do not have the disease of secondary hyperparathyroidism. However, some individuals with very low PTH levels have been treated for secondary hyperparathyroidism effectively and perhaps overly treated and their PTH level suppressed in response to pharmacological interventions. Under these circumstances for safety reasons treatment reductions or withdrawal would be considered appropriate. The primary concern here relates to issues of fracture risk, potential for vascular calcification although the evidence supporting those adverse outcomes is somewhat tenuous.

Thank you.

DR. PACE: Okay. Lisa?

Okay. So we'll go back to our process of having the person introduce the measure and give a summary of the preliminary vals and raise any issues, and we'll do it criterion by criterion. So we'll start with impact.

DR. LATTS: Right. And first of all, I want to thank Amgen for actually two very well written proposals. But I thought that the proposals were very well written. And I want to thank NQF for assigning them to me when I had to review things that I'd most happily forgotten since medical school and I'm definitely going to need help from my nephrology colleague in terms of the parathyroid calcium phosphorus access.

So, the Amgen rep said, this two proposals are regarding the use of vitamin d analogs and calcimimetics for high an low parathyroid levels. Instead of an overall, we'll go through measure by measure.

DR. PACE: Let's do measure by measure and correct as we need to.

DR. LATTS: So the first measure then, 1655 ESRD patients with parathyroid greater then 400 who are not treated with calcimimetic or vitamin D analog. First looking at importance and impact, fairly good

agreement among the reviewers that this is something that is moderately to high importance, important to measure. Impact, yes.

DR. PACE: Any comments on impact or are you ready to vote on that? Okay. Let's vote.

CO-CHAIR CROOKS: Vote.

MS. RICHIE: And Lorien, impact?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Okay. The results are eight votes for high, 12 for moderate and one low.

So, performance gap.

DR. LATTS: Okay. So again, between the reviewers and within the document they have review from a large dialysis organizations and from the Dialysis Outcomes and Practice Study showing fairly significant variation, I thought, between patients and between facilities.

So within patients in the large

dialysis facilities, 16 percent of patients would have been tested positive for this measure and 25 percent in the DOPPS study. Within facilities, 39 percent and 42 percent respectively would have tested positive on this measure.

CO-CHAIR CROOKS: Okay. Other comments regarding performance gap. Okay. I think we're ready to vote on that point. So let's vote high, moderate, low or insufficient.

MS. RICHIE: Lorien, performance gap? Lorien?

DR. DALRYMPLE: Oh, I'm sorry. Moderate.

MS. RICHIE: Thank you.

CO-CHAIR CROOKS: The results: 8 votes high, 13 moderate.

Now onto the body of evidence.

DR. LATTS: Right. Quantity. So in quantity of studies, there were nine publications reviewing 15 studies looking at

the relationship of moderate to severe hyperparathyroidism associated with bone disease and risk of death. I think this is where we start to get controversial, and I think this will be a very engaged discussion. And, you know again, we'll refer to some of my nephrology colleagues as to the evidence. But I think that there appears to be a good link between the relationship of parathyroidism to bone disease. From there on, it gets a little fuzzier and again, would like some of my esteemed colleagues to weigh in.

DR. KLIGER: Okay. So I'll weigh it in.

The data, I think are pretty clear about a correlation between the presence of PTH levels and poor outcomes. I haven't seen any data, though, suggesting that altering that levels affects outcomes.

DR. NALLY: I'm sorry. Mine went on first.

Okay. I'll get myself in trouble.

So there are no randomized controlled studies affecting that outcome. And as stated, the measure includes monitoring whether or not patients are on two classes of agents; vitamin D analogs or another treatment. And that implies that that is the right thing to do, but there's no randomized control trial data there. So to me that's the conundrum. And then when KDIGO looked at this and then there was a commentary, a U.S. commentary their conclusion which we talked about in great detail in January, was that issues related to control of phosphorus and PTH did not appear to meet a standard for performance measures. So to me that is my concern with incorporating drugs into this measure related to the monitoring of PTH.

DR. LATTS: The question I have, and the authors point this out in the performance metric brief, is that is this a randomized control trial that could be done, or would it be unacceptable to have a high PTH

that remains untreated under today's practice standards?

DR. BERNS: Well, I think it could and should be done. I think that Joe makes an important point, KDIGO did not feel that this should be a performance measure. And it also looks at PTH in isolation when really metabolic bone disease management is what is a calcium, what is the phosphorus, what is the PTH, what have been the trends in those over time, as opposed to looking at only one laboratory value at one point in time in insulation I think is actually bad care.

DR. FISCHER: And particularly with the variability in PTH. There was a study that showed you have to check it -- I may get this wrong -- but in double digits the number of times you have to check it before you have a stable value. And I'm sure anecdotally many of the people here around the table in their own unit have rechecked PTH values and it's 600, and then it's 200. And I think that's

why this study was done that showed the remarkable variability and regardless of which assay you were doing; they looked at different ones. But then I think the second thing I'd just add is then what is the appropriate threshold? I think this gets into the variability in the assay. Here it's greater then 400, I don't know how great the evidence is for that and particularly in the backdrop of a very fickle assay I think that's very problematic.

DR. KLIGER: The developer mentioned a safety signal. So I think we also -- I want to make sure we're clear about that. Because my interpretation is that we need to consider a safety signal at the low end where we might have prescription of medicines where there's no indication for it. And I'll just ask the developer just to clarify. He mentioned safety; are you concerned about safety at the low end or is there any evidence of a safety concern at the high end?

DR. GOODMAN: I think we're concerned definitely at the high end. Again, our view is -- and there's evidence I think that is compelling that this is a progressive disease. Once the process of parathyroid gland hyperplasia becomes established, it's a progressive disease. And I think KDIGO actually acknowledges that in recommending that if there is biochemical evidence of progression, then an intervention to control that progression and to prevent values from reaching levels that are associated with extreme risk is appropriate.

With respect to the PTH assay measurements, granted there are many commercial assays available and they provide numerically different results. They are, however, all marketed under FDA scrutiny and they satisfy the criteria the FDA establishes for marketed diagnostic products. So it's important for providers as well as clinicians to understand which assay is being used and

how it relates to previous assays considered to be gold standard. But the reliability of these is greater than is generally discussed.

Looking at any of these observational studies, or looking at population data in a population receiving many different treatments and 80 to 85 percent of this population are on a variety of treatments, short term changes in PTH are readily understandable. We've looked at this in datasets in individuals with untreated disease. So they're not confounded by concurrent treatment with either vitamin D or a calcimimetic. And if one looks at individuals with values above 400 off treatment, then looks retrospectively over six or 12 months to document that they've received no treatment, the interval change over that six or 12 month period is in the range of 40 to 50 percent in terms of their PTH level.

And if one looks at two consecutive measurements separated by three months, two-

thirds to three quarters of the time the second measurement is higher than the one obtained three months previously.

So I think there's good evidence in individuals who are not treated that this is a progressive disease.

So to your point, Alan, I think that there is risk at the high side in terms of disease progression.

CO-CHAIR CROOKS: Jeff?

DR. BERNS: Bill, do you have information available about bone disease itself or in these patients or sort of at these different PTH levels rather than just the PTH level? In other words, bone biopsy data?

DR. GOODMAN: We've just last week looked at data from a study that we undertook as a post-marketing commitment in Europe. And it is pretty clear that patients with PTH levels above 500 to 600, the overwhelming majority of them have evidence of

hyperparathyroid bone disease as documented by bone biopsy.

DR. NALLY: And were those patients naive to vitamin D analogs and calcimimetics?

DR. GOODMAN: About half of them had previously been treated with a vitamin D analog. Very few had been previously treated with a calcimimetic.

CO-CHAIR CROOKS: You know, I think it's clear that a good nephrologist is going to address a high PTH level as part of their care. And the issue I think we're grappling with is without good evidence that an intervention makes a difference in key outcomes, is this something that should be a National Quality Forum voluntary consensus standard?

Before we start voting on the body of evidence questions, is there anymore discussion?

DR. GOODMAN: If I could just add one more comment to address Jeff's point.

CO-CHAIR CROOKS: One more.

DR. GOODMAN: We certainly are not advocating looking at PTH in isolation, but it is an independent measure of the disease. There is no other parameter that can be measured other than bone pathology to inform about this disease. So calcium or phosphorus levels per se will not provide any diagnostic information whatsoever with respect to the presence, absence or severity of secondary hyperparathyroidism.

CO-CHAIR CROOKS: Okay. Lisa, any other?

DR. LATTS: No. You know, I find myself struggling with some of the evaluations for putting the discussion we just had in context with the NQF sort of structure in that, you know obviously there was a very robust body of evidence presented. It's just not directly on the question. So I think that's sort of the key thing to consider as we are voting.

CO-CHAIR CROOKS: Okay. Are we ready to vote, first on the quantity of studies? You've seen the chart, five or more is high, two to four moderate, one would be low. Let's vote.

MS. RICHIE: Lorien, quantity?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Okay. The voting: Four high, 11 moderate, two low and 4 insufficient evidence. Okay.

The next is the quality. High, moderate -- are we ready to vote? Any other discussion here? Okay. Let's go ahead and vote.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: All right. We have one high, seven moderate, eight low and five insufficient evidence.

Let's go ahead and vote on consistency results across the body of evidence. High, moderate, low or

insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: That's 21. We have nine moderate, six low and six insufficient. So applying that to our algorithm, I think this would fit the third row, right? Quantity medium to high, quality low, consistency medium to high. So this would pass if the potential benefits to patients clearly outweighs potential harms.

DR. PACE: No, I think the --

CO-CHAIR CROOKS: Did I get that wrong?

DR. KLIGER: I'm not sure I agree with that assessment. If you look at the consistency --

DR. PACE: Right.

DR. KLIGER: -- low end cannot determine for the majority

DR. PACE: Right. So --

CO-CHAIR CROOKS: So what was the

consistency?

DR. PACE: Can you go back to the results for consistency? Please display it again. Yes. So it was low is insufficient, 12.

CO-CHAIR CROOKS: Okay. So we have to give that low. The insufficient's hard to figure how that should count, right?

DR. PACE: Right. Well, insufficient mean you really can't rate it. And I think we have to combine that with low versus just compare low to moderate.

CO-CHAIR CROOKS: Yes. Okay. It feels that way. I'm not sure it means that. Because it may that if they're saying if I had that insufficient evidence, I might feel it's good. Okay. So we're going rate this as a low. I think --

DR. PACE: No. Not passing evidence.

CO-CHAIR CROOKS: Well, then going back to the chart, go to the next -- so then

we would be down to the fourth line, the fourth row, correct? Everyone agree? Okay.

DR. PACE: So any concerns about that? Because we can rediscuss if needed. So basically what we're saying is this would stop here because it didn't pass evidence. All right.

CO-CHAIR CROOKS: Okay. So let's go to the next measure. 1658.

DR. LATTS: Sorry. This is the flip side, overuse measure looking at whether someone with a low PH -- or low PTH below a certain threshold, and that threshold has been chosen as 130, is being treated with a vitamin D analog or a calcimimetic.

In terms of the reviewers, the initial importance was sort of all over the place with three mediums, one high and two lows. So definitely all over the place, although I would change my high to a medium after this discussion now.

And, you know I think our previous

discussion is still very valid in terms of -- I would assume there is not a lot of information also on the flip side of what to do with a very low PTH and the validity of improving that number by stopping one of these drugs, as well as the variation in the lab tests.

DR. PACE: So let's focus on impact first.

DR. LATTS: Okay.

DR. PACE: And see what the other reviewers wanted to say.

DR. FISCHER: Really, I misunderstood impact before coming. So I would change my low up there to a moderate. Because I was focusing very narrowly on the impact of this. Karen kind of elaborated, that's more of the broader impact of the topic area. So, with that new knowledge I would change my vote.

DR. PACE: All right. Impact, high, moderate, low, insufficient.

MS. RICHIE: And Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: One high, 20 moderate.

So going onto the performance.

DR. NARVA: I'm just curious. In the application was there any data maybe from Part D or from someplace to suggest how big a problem this is? Where there's simultaneous PTHs and drug utilization?

DR. LATTS: Well, funny you should ask that. That's the next one, performance gaps.

CO-CHAIR CROOKS: That's where we're going.

DR. LATTS: Yes, that's where we're going right now. So the same two databases were used as for the last one, a large dialysis organization using their electronic medical records, and then the DOPPS study. And in this then looking at low PTH still treated, they found a 60 percent of patients

in the large dialysis facility, 46 percent of patients in the DOPPS study with serum PTH values less than 130 still treated with a vitamin D analog or a calcimimetic. And then on the facility side, 59 percent of the facilities and 58 percent -- I'm sorry. Fifty-nine percent of the large dialysis organization facilities, 58 percent of the DOPPS study facilities had patients with a PTH less then 130 still being treated.

So, fairly large numbers that would test "positive" for this measure.

CO-CHAIR CROOKS: I found that pretty persuasive. And also thinking of this as a safety metric, you know, that that's kind of alarming. We'd have to look deeper to really know exactly what's going on with those, but I found that persuasive.

DR. BERNS: The only comment I'd make is that's pretty old data at this point. That's from 2007, I think all of it if not most of it. So for whatever it's worth it's

rather outdated at this point.

DR. FISCHER: And I have a question just for clarification maybe from the steward. But one concern I had is this treats the treatment decision kind of dichotomous. Either you're giving treatment or not. And I guess one of the concerns, I'm sure others shared this, is what if the provider had made a substantial dose reduction in the vitamin D analog or the calcimimetic? And this may be a limitation of the secondary data sources they were using, and then it also I think has concerns just for how this is written. But I wanted to make sure I understood from them that, I guess, that that wasn't available and/or is that something that they were meant to incorporate in the way this is written?

DR. PACE: Are you talking about access to, like, over-the-counter?

DR. FISCHER: Yes. No, no. In other words if this treats you, either you were being treated with a vitamin D analog or

a calcimimetic or not, in a case like this what if the provider had made a substantial dose reduction in the medication?

DR. LATTS: And I've had that exact same issue.

DR. FISCHER: Yes.

DR. LATTS: So there's a three month window we're looking at. You get the lab value, the provider makes a change, either stops or massively reduces the drug, and you would still test positive because they were on drug during that three month window.

So, I think, you know, for us to -- there would need to be an opportunity for the provider to get "credit" for making the change.

DR. GOODMAN: Yes. Certainly again you'd have to engage with the trending over time, sequential measurements. But at these levels these are considered to be very low among patients undergoing dialysis. And so continuation of treatment here, you know,

after dosage adjustment, you know, would really be considered over-aggressive of both therapeutic agents.

DR. LATTS: But if within that 90 day window that you're looking at for the measure, someone is on-drug, gets their treatment results, stops the drug; because they were on-drug within that 90 day window -- it's not 90 days after the positive test result was my reading of the measure. You get that test result -- it could be that you get the test results in the last month of that 90 days, you were on-drug up until that test result and then stopped it, and you would still test positive. Unless I am misreading the -- and that's sort of getting into the Part 2 in terms of the reliability. But unless I'm misreading it, that's how I'm taking it.

DR. GOODMAN: Now, granted, there may be some refinement that needs to be done there for sequential testing.

CO-CHAIR CROOKS: All right. So are we ready to vote on the performance gap? Any other questions? Okay. Let's vote.

Wait, wait. Back up. Did we already vote on -- no. Okay. Here we go.

It's been a long one and a half days.

MS. RICHIE: Lorien, performance gap?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Someone out of the room? Let's go with 20. All right. Four voted high, 15 moderate and one low. Okay.

So onto the body of evidence. This is, yes, not an outcome. Right.

DR. LATTS: So there were 12 studies that were reviewed to look at the parathyroid hormone over suppression in renal disease. It seemed a little more on point to me then the last set, perhaps.

DR. PACE: So we'll talk about the quantity, quality and consistency and then go

back and vote.

DR. FISCHER: I mean, I just have similar concerns with the last measure where: one, you have other parameters of bone metabolism that vary over time and are highly time dependent, and this takes one and kind of takes it in a prescribed time window. So I think those are important things in decision-making in trying to assess what's the best treatment strategy.

And then the second thing is is the exact threshold. Once again, we have these defined thresholds, here it's less then 130. How strong is the evidence for that particular cutoff, particularly taking into account the other comments that others have made about the variability, even within any given assay you do for PTH, and not having other bone metabolism parameters as part of kind of the gestalt of the overall impression of a patient's parathyroid disease.

DR. KLIGER: Mike, I just heard

comments about the quantity, and I haven't heard the Steering Committee's thoughts about the quality yet. So can we just vote on the quantity and then hear your quality comments.

CO-CHAIR CROOKS: Yes, I think that's fine.

DR. PACE: Okay.

CO-CHAIR CROOKS: They mention that 12 studies were involved in the body of evidence.

Okay. Let's go ahead and vote. High, moderate, low, insufficient.

MS. RICHIE: Lorien, quantity?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Okay. Eleven voted high, eight moderate, one low and one insufficient.

So, to the quality.

DR. LATTS: And again we'll ask my Committee members here to help me weigh in on the quality.

The studies I think were a little

more on point as to the relationship between parathyroid hormone and the morbidity associated with bone disease, cardiovascular disease, et cetera. No direct link to mortality.

And then they also, and I'd again like my Committee members to help me, the Palmer study looking at a sort of pseudo meta-analysis looking at 14 cohort studies assessing the quality of evidence for the association between phos, PTH, calcium, risk of death and cardiovascular mortality. And there was not a tight relationship found in that study. And that review it didn't met the criterion of meta-analysis, but in that cohort review.

So the authors felt that this was directly on point and there were some problems with this analysis.

DR. NALLY: I have a fundamental struggle here, given the concerns at different levels about the evidence and the absence of

black-and-white evidence. But on the other side, I do tend to view this as a safety monitoring issue. And if one has a very suppressed PTH on vitamin D analogs or calcimimetics, I think most people in the room would want to remove those drugs, maybe without the most profound evidence in the world, but I think we think that's the right thing to do.

But again, the concern with the measure as written is just what Lisa articulated. You might have drastically reduced or, hopefully, stopped but the way the measure is written, because of this 90 day window business, it may be perceived that the patient on-drug -- yes, the patient was on-drug when his PTH was 300, but now you get the number back and it's 100 and you're going to stop it tomorrow or today.

In my heart of hearts I believe it's an important safety measure that we should consider, but otherwise there's a lot

of flaws in the evidence per se.

DR. LATTS: And maybe what I'd suggest is maybe let's vote on -- because I think you're talking about reliability. And I think we can fix it. If we want to proceed with the measure, I have some thoughts on how we could fix the measure to get to that in a more direct fashion. Because you're right, as written it's not appropriate, I believe.

CO-CHAIR CROOKS: Well, as a reviewer I had the same dilemma that Joe's describing. I don't think the quality of the evidence is sufficient, yet I agree that this is important in the sense that it's a safety measure and -- does it rise to the level of needing a National Quality Forum standard? You know, that's what I'm debating in my own mind.

I'm wondering, this could be one of those measures where we say the quality isn't there but maybe the benefit exceeds the harm.

Alan?

DR. KLIGER: Just a quick clarification. The evidence shows the morbidity of the low levels. The measure has to do with stopping the drug. Do any of the studies deal with stopping or not stopping the drug?

DR. FISCHER: This is quite well written. I mean, on page 13 their last paragraph kind of states exactly that, that the overall quality of evidence -- according to this ,there's guidelines, but they say that it's not clear what to do or -- evidence about a level, a consensus about evidence PTH value which would trigger an action of any kind, whether we stop or dose reduce. Or what action should be dose reduction versus dose continuation is not very well known. And then they kind of go back to citing some things in the guidelines, which I think are more of a product of expert opinion again.

So, I think that it's quite well written and put together. And I think it

underscores that there's a lack of evidence and there's uncertainty. But there is expert opinion floating around. And I guess then I think one has to weigh that expert opinion and lack of hard evidence versus the safety concerns that others have mentioned.

DR. LATTS: I actually think that, you know, when you guys are looking for an example and a really well-written review to give to potential measure developers, this would be a good example. It really is quite well-written.

DR. NALLY: But the conundrum here is that it is actually so well written that that paragraph that was alluded to I think strikes it down and it seems to be the right thing to do, but we don't have clear-cut evidence. So it might be a clinical guideline, but it maybe should not be a performance measure.

DR. LATTS: Well, you know, one of the things we have actually not talked about

in this meeting, although I remember discussing it back in January, is that this an untested measure. So it would only if we endorsed it -- and you know again, I think there would need to be fixes first -- oh, there's no time limit anymore. Okay. Never mind.

DR. PACE: And let me just remind you, too, because I think Peter mentioned it. But in this situation, as you're talking about, even though it might pass evidence, if you really think this measure is a safety concern and, as Peter said, the benefits greatly outweigh the harm, then you can proceed on that basis. So, I just want to be sure that you're aware.

CO-CHAIR CROOKS: Yes, if the voting goes low for quality but moderate to high for consistency, then we have the option of saying, yes, without even doing anything extraordinary --

DR. PACE: Right.

CO-CHAIR CROOKS: -- we can just say that the benefit outweighs the harm is the next question that comes up.

Okay. Thank you. Any other discussion before we vote on quality on the body of evidence? Okay. Let's vote.

MS. RICHIE: Lorien, quality?

DR. DALRYMPLE: Low.

CO-CHAIR CROOKS: Okay. The lows have it. We have one high, three moderate, 14 low and three insufficient. Okay.

So let's go on to the consistency question. Any discussion about consistency? Okay. Let's vote.

MS. RICHIE: And, Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: All right. So we have two voting high, ten moderate, four low and five insufficient. So I think we can give this a moderate? Do you agree?

DR. PACE: Yes. I mean, it would be -- right.

CO-CHAIR CROOKS: So that gets us to row three where we have moderate or high quantity, low quality and moderate to high consistency. So now we can consider the question if the potential benefits outweigh the harms, we can vote yes and it would pass the evidence review. Discussion? Okay. Can we vote then?

DR. PACE: Okay. We can vote. Go ahead and vote here. That's fine. So yes, if the benefits outweigh the harms.

CO-CHAIR CROOKS: Yes.

DR. PACE: Right. This is actually if it hadn't passed evidence at all. So I think we could actually stop unless someone objects to that conclusion. Or do you want to go ahead and vote on it? That's fine.

CO-CHAIR CROOKS: Are there people in the Committee who would argue that the potential harm outweighs the benefit of stopping the drug when the PTH level is low? No. Okay. So I think we can just say that it

passes on that.

DR. PACE: And we'll give that rationale.

CO-CHAIR CROOKS: Okay. So we can move on to --

DR. PACE: Reliability.

CO-CHAIR CROOKS: -- reliability and validity.

DR. LATTS: So the numerator here is the number of patients from the denominator with PTH less then 130 who continue to be treated with a calcimimetic agent or a vitamin D analog. There's a three month reporting window. The denominator is anyone who is hemodialysis or PD 18 years or age or older, been in the facility for 30 days who have been on dialysis for better than 90 days.

We've talked previously about some of the issues with this in terms of anytime within that 90 day window, is my understanding, if you have a PTH less then 130 and if anytime within that 90 day window

you're treated with a vitamin D analog or a calcimimetic you are in the testing yes category. So there's no sort of sequential time issue, which again I think could potentially would be fixed with some fixes to the sort of -- you could use an index event of the PTH and then look for a 90 day window after that or, your know, have some sort of -- I think this could be fixed if we decided it's important to proceed. But I think as currently written it is not testing what you want to test, which is does the facility and/or physician or clinician appropriately make a change to therapy as a result of the test.

DR. KLIGER: So, Lisa, if that's right, and I think you're right, do we try to fix it now or do we vote on the flawed current measure?

DR. PACE: I think what we've learned is it's best to vote on the measure as it is and then if someone wants to try to

suggest a change, that way we'll know where we're at better. And also, I'm sorry if I missed it, did you talk about reliability testing as well besides the specification?

DR. LATTS: I did not. So reliability testing was not done -- or was it -- yes, validity testing was not done. Thank you.

Yes. I'm sorry. Yes. So they used the large vast organization with the EHR, you know and again we have all the issues we discussed yesterday using EHR data, and some of the issues there.

DR. PACE: Okay. So they were invoking that they were doing data element validity testing --

DR. LATTS: Yes.

DR. PACE: -- and then we allowed them to skip reliability. So we'll address that under validity.

Ruben?

DR. VELEZ: I would like to add

here, if possible, maybe on the part of the steward, to think about an exclusion. Some networks are beginning to see more parathyroidectomies. Those patients will need vitamin D analogs initially to maintain calciums and they will have a low PTH. So I think we should think about that exclusion.

CO-CHAIR CROOKS: Say that again, Ruben. I didn't follow -- which group are you thinking about excluding?

DR. VELEZ: Patients that a recent parathyroidectomy that need to be on vitamin D analogs.

CO-CHAIR CROOKS: Under specifications, I was maybe sort of overlapping the feasibility a bit, but I'd like to ask the developers. You mentioned the data source could be CROWNWeb data. Have you worked out an agreement with CMS? You know, if this is passed, who is actually going to be doing the data, where does the data come from? Does Amgen do the calculations and where will

you get the data?

MR. NUSBICKEL: Yes. We had a very brief email conversation with Tom Dudley. And we suggested that in the data field which they currently have in CROWNWeb where they collect vitamin D that they also collect calcimimetics.

We also indicated that it would be necessary for them to provide the conversion, you know, given specification on which assay was used at each of those facilities.

And so we've just had the initial conversations so there's no agreement in place.

CO-CHAIR CROOKS: Okay. So basically you would make this available to CMS to use, otherwise it wouldn't otherwise be probably used, is that right?

DR. PACE: And if any NQF endorsed measure can be used by anyone.

CO-CHAIR CROOKS: Right. Although, they have to go to the measure steward to make

sure that they're doing it right, in general?

DR. PACE: Yes. It would be the endorsed measure.

Are you ready to vote on reliability which includes specifications on the measure as it is? And then if it doesn't pass here, you can bring up if someone wants to propose a modification, you can do that?

CO-CHAIR CROOKS: Okay. Are we ready to vote on reliability? Okay. Let's do.

MS. RICHIE: Lorien, reliability?

DR. DALRYMPLE: Low.

CO-CHAIR CROOKS: That's 21. One high, 3 moderates, 16 low, 1 insufficient.

DR. PACE: Correct.

CO-CHAIR CROOKS: Will somebody in the majority explain to me why they're feeling reliability is low?

DR. KLIGER: The specification issue that we've discussed.

CO-CHAIR CROOKS: That I --

DR. KLIGER: No. Lisa was the main proponent.

DR. LATTS: That the measure as specified does not look at whether somebody appropriately responded to a low PTH level.

CO-CHAIR CROOKS: Okay. Thank you.

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

So it was really the specification not the reliability issue. Okay. Thank you.

All right. So do we stop here or do we move on? Because this is kind of a deal killer at this point.

DR. PACE: This would be a deal killer. So the question is whether someone wants to propose a modification to fix the specifications and then we could vote on it.

DR. FISCHER: I thought we were voting on it. I thought the last vote was voting on reliability as is and it included specifications. I just want to make sure I voted on what I thought I just voted on.

CO-CHAIR CROOKS: That's right. They're kind of bunched together.

DR. PACE: Right. You're right.

And maybe let's do this so we don't confuse things. Let's go ahead and vote on validity as well. And then we can talk about potential modifications if someone wants to bring that up. That way we won't get confused of where we're at.

CO-CHAIR CROOKS: Okay. Lisa, so how was validity demonstrated?

DR. LATTS: Okay. So validity was demonstrated using testing from this large dialysis organization using data on 43,000 patients. They looked at this database and also 81 facilities from the DOPPS data. They found that -- let me look. Basically the data showed that they could get the measures out of the datasets, and again this was EMR data so it was not CROWNWeb data so it's a little different from the sets we've had currently.

There is the issue that the

developers mentioned just a minute ago that the calcimimetic is not in the CROWNWeb database. So they have asked CMS and CMS has apparently agreed via this email conversation to instruct facilities to use the vitamin D analog element if the patient is on either a calcimimetic or a vitamin D analog. So, there's a little bit of an issue there.

There also is the issue that was mentioned in the last -- actually, we didn't get to it in the last one. And you guys again might know a lot more about this than I do, there's some problems with the PTH tests in that there's no comparability across testing. So the reference range from one test is not comparable to the reference range in another test. So there are calculations that have to be done to normalize the ratio between testing which seems to me to be quite a nightmare and I think causes some significant problems in how these tests would be interpretable. Because you can't use 130 as an absolute

cutoff. There needs to be some machinations depending on what test your particular reference lab is using to translate that into 130. So I see that while there are calculations that can be made to normalize it, I see this as a bit of an issue and a problem.

DR. PACE: I'd just like to clarify one thing. Even though we've talked that they were trained to address data element validity, they really didn't get at data element validity. They had aggregate numbers that they compared to study data. So we still don't necessarily know that --

DR. LATTS: It's not been tested in its form, yes. It's the elements tested via the scientific databases, the research databases

DR. PACE: And it's at a very high level, so we don't really know what the data element --

DR. BERNS: So just to clarify there, CMS -- there's no reporting right now

from dialysis facilities to CMS for any of the vitamin Ds, is that correct? The facilities don't currently report vitamin D use, oral, intravenous or calcimimetic use. So there's no way to know whether--

DR. GOODMAN: No. Not currently. Only billing data.

(Simultaneous speaking.)

DR. PACE: Okay. Wait. And what about the PTH level, is that being reported?

DR. GOODMAN: Not currently.

DR. PACE: Okay.

DR. LATTS: But that's clinically enhanced data. You should be able to get that through the lab vendors. Not easy, but possible.

DR. PACE: So the question right now is we're talking about a specific measure that's been put before us using a particular data element and did they demonstrate validity of the data or of the score that will be used for the measure that's being presented?

DR. KLIGER: Right. And I guess that's what I was going to ask again, Karen. Because as you're the expert who understands the mechanism of validity testing. And I understood your comment to be that the elements were not there to test validity. So we don't have any information on validity, is that correct?

DR. PACE: It seems that way to me from looking at what they provided. And maybe we can pull that up in the application, the 2.B.2.

DR. NALLY: I think the interpretation currently is insufficient would be --

CO-CHAIR CROOKS: I think they're trying to make the case that if they have the data, it would be valid. You know, as I'm reading it, they --

DR. PACE: Well, I think they have data from the LDOs, as Lisa was saying.

CO-CHAIR CROOKS: Right.

DR. PACE: And they looked at -- so they --

CO-CHAIR CROOKS: And they compared it to DOPPS data.

DR. PACE: So they looked at kind of aggregate numbers and then said well this is similar to what's in the DOPPS database. But it's not specifically looking at the data for this patient compared to some authoritative source of the data for that patient. So that's the point I'm making in terms of what does that show when you're--

DR. FISCHER: But I thought the idea is that DOPPS is kind of the gold standard because DOPPS is a prospectively controlled study, right, where you had research assistants asking patients and writing down their medications. So I guess I thought the idea was is that they were showing that the data we were able to extract from an LDO correlated highly with DOPPS data, which is a gold standard in terms of --

DR. PACE: Right. But it is the same facilities and same patients? See, that's the question.

DR. FISCHER: No, no, and that's absolutely -- no, it's not. Because DOPPS is, right, a worldwide study on several different continents and this is from LDO in the United States. No, it's not the same patients. So it's an indirect -- I'm just thinking that there were other examples that we've talked about here today and yesterday where there was an indirect way to try to use correlation with samples that are not exactly the same in an attempt to demonstrate validity.

CO-CHAIR CROOKS: They did say they used U.S. DOPPS and use worldwide DOPPS.

DR. FISCHER: I overlooked that.

CO-CHAIR CROOKS: And so an LDO, you know the two big LDOs are national companies and you would expect the DOPPS and their population should be very similar.

DR. PACE: And do we have that up,

2.B.2, the results? The Tables 9A and 9B, can you bring those up?

CO-CHAIR CROOKS: Well, Table 8 is the first part of the results and then 9A and 9B is the second. There's actually two tests that are -- Bill, you're invited to explain, or one of you, the validity testing.

DR. GOODMAN: Well, I mean the data that were used here are essentially equivalent to the kinds of data that would be reported to CMS or to CROWNWeb.

DR. PACE: But this is basically population level. It's not even at the facility level, right?

DR. GOODMAN: Correct.

DR. PACE: So -- okay. So I think you all can weigh that, as Michael was saying, but we're just pointing out you have to know what it is and isn't telling you.

CO-CHAIR CROOKS: Okay. So are we ready to vote on validity? Any other discussion? Okay. Let's go ahead and vote.

MS. RICHIE: Lorien, validity?

DR. DALRYMPLE: Insufficient.

MS. RICHIE: Thank you.

CO-CHAIR CROOKS: Some may have voted too soon because it took a while for it to come up. So you might want to vote again. Here we go. Okay. We're three moderate, six low and 12 insufficient. Okay.

So we have problems with it, both specification and validity. So short of getting CROWNWeb going, which they can't do immediately.

DR. LATTS: Yes. I mean I think even we fix the reliability issues which we might be able to fix, we have the validity issue. And I just think it might not be ready for prime time this round.

DR. PACE: But I mean unless someone has a suggestion that -- I mean, so we have a couple of things here, but one kind of impacts the other.

So we do accept face validity, and

that's something that they could address in a relatively short time. But that means then that they would have to do something about reliability testing. And I don't know if--

CO-CHAIR CROOKS: Specifications.

DR. PACE: And also, definitely, the specifications.

So I don't know how strongly the Committee feels about asking the developer to think about these things rather than proposing -- Joe?

DR. NALLY: Just a point of clarification about existing endorsed measures. Is there any endorsed measure in ESRD related to PTH monitoring without these drugs involved? In other words, there's no measure that looks simply at a low PTH, correct? Thank you.

CO-CHAIR CROOKS: Kathleen.

MS. LeBEAU: I might just remind everybody that -- way the conversation was that this is a safety issue and that this is

an evolving process. So, it might be worth our time to see what we could do to make this -- you know, address the deficits.

DR. KLIGER: Yes. Yes. I agree with Kathleen. Rather than drop it, my advice would be that we go on with this with the recommendations of validity and specification testing as we've discussed for the developers to give us.

CO-CHAIR CROOKS: Fortunately or unfortunately it's clear that our work isn't going to be done today and that there would be a several week period of time for them to address some of these specific concerns.

DR. PACE: So do you want to take a few minutes to talk about what the specification changes you're thinking would be useful so that we can give them that input? And then we will follow-up with them about how we can address the other aspects?

DR. LATTS: I mean, my suggestion I'm definitely open to helping refine this

would be to use the PTH level as an index event and then look at the 30 or 60 days after that event for prescriptions of a vitamin D analog and calcimimetic to give the facility and the clinician time to effect change after the results are obtained.

You know, it's obviously a more complicated measure. You would have to exclude folks that had a subsequent PTH that was above that range that were then restarted. So there would have to be some machinations. But I think it could be done.

DR. BERNS: And the other suggestion might be to look at this, again it'd be complicated, but use the cutoff value of two times the upper limit of normal for that lab rather than a specific number.

I think the recommendation from KDIGO and others reflecting the variability of the assays or differences between the assays is that rather than 130, the appropriate number might be two times the upper limit of

normal for that lab's assay.

CO-CHAIR CROOKS: So that's something for consideration.

DR. FISCHER: And then the other specification -- was there consensus that all vitamin D analogs and calcimimetics should be stopped or is it the idea that stopping one or the other, if someone's on both or a dose reduction if they're on one is reasonable in terms of -- I mean, I guess that's one other thing that I have a little bit of trouble with that it's kind of written once again binary, dichotomous; everything is stopped or not.

DR. LATTS: Well, what I was wondering is when we did the hypertension measure yesterday -- was it just yesterday, there was a plan, a treatment plan. And could it be something like that where there's a treatment plan to address the low PTH?

DR. PACE: I'll just say that those are even more complicated.

DR. LATTS: I know, I know.

CO-CHAIR CROOKS: Yes.

(Simultaneous speaking.)

DR. LATTS: I know that's why I was sort of hesitant even to mention it.

DR. PACE: But I guess the other question, because we in the last project we had the kind of safety measure for the hypercalcemia, I believe. And it was just the level and not associated with drugs. So my question to you is would that make sense in this respect?

DR. KLIGER: This is different.

DR. PACE: Okay.

DR. FENVES: And if I may comment on -- I completely agree with Michael's comment because one size doesn't fit all. This is a complex -- I mean it's so patient dependent depending on other factors on what you might do. It would be not good to mandate, let's say, or assume that we mandate stoppage of those.

DR. PACE: Jerry?

DR. JACKSON: Just a point of clarification. Since this specifies vitamin D analogs, should the patient have a low 25-hydroxyvitamin D it would not be preclude them being on vitamin D itself, vitamin D3, is that correct?

CO-CHAIR CROOKS: Yes. This doesn't address vitamin D3, right, Bill?

DR. GOODMAN: Right. We're specific of vitamin D analogs, not native or nutritional vitamin D.

CO-CHAIR CROOKS: And the other issue about validity is to consider making the case as face validity addressing the appropriate related issues on that instead of this type of validity.

Ruben?

DR. VELEZ: Just remind the possible exclusion that we mentioned earlier.

CO-CHAIR CROOKS: For post-parathyroidectomy patients that should be an exclusion.

DR. FISCHER: And one of the things is that there may be a limit because when you talk about provider actions, particularly when it's dosage reductions, similar with blood pressure, I think this becomes very challenging. Because it becomes complicated, as Karen mentions. Not only to get to kind of right on algorithm, but then to actually have data such as that.

So let's say you were able to write something where it was a dose reduction, how are you going to go to CROWN data or somewhere and be able to figure that out, you know be able to establish that change in action over time? And this gets, I guess, now to feasibility and I don't want to start muddling issues. But just as we're talking about responses back to the steward, I think correcting one thing may lead to difficulties elsewhere down the road.

DR. PACE: The other thing I think to think about is that, you know from

performance measurement standpoint you can't expect to have a standardized measure that will encompass every exception. And so the question to you all is so if it's left as is with expecting, you know kind of the on/off is that going to in a variable way effect scores of facilities? I mean, are patients going to be kind of -- you know, is it a random occurrence? Is it a big issue? I mean that's the other thing is that if it's a small minority of patients, then it's not going to effect overall performance scores. And we don't have to expect 100 percent or zero percent on this kind of measure. But if it's something that's variable across facilities? You know, so we have to kind of think about that, too.

CO-CHAIR CROOKS: Yes. It may be that zero isn't the right percentage. Ten percent may be correct, you know. And so you can compare -- it's a facility measure, so if one facility is 50 percent and the rest are at

ten percent, then you have an issue. But if there was 12, 13, 8, 9, that's not a significant variation.

DR. LATTS: Well and I wonder if we could use persistency to help us in a sense that what if we were to do something like two elevated -- sorry -- suppressed PTH levels in subsequent months, in that case would it be much clearer that the drug should be stopped as opposed to just reduced?

MR. McMURRAY: Peter, it seems to me that with all the discussion we've had here today to try figure out how to fix this in this meeting doesn't make any sense. It would seem to me that either this needs to go and come back in a different form with more thought, or there needs to be a group put together to kind of think through this with the contractor to make this happen.

We could sit here and debate this all day.

CO-CHAIR CROOKS: You're exactly

right. What we've done I think is offered some advice to the developer, issues that are of concern to the Steering Committee and offer them a short time window to redress this, if they wish to. And that's all we can do at this point.

Thank you.

Okay. With that sage advice from Stephen, let's take a ten minute break. We'll resume at 20 minutes to.

(Whereupon, the above-entitled matter went off the record at 10:29 a.m. and resumed at 10:47 a.m.)

CO-CHAIR CROOKS: Okay. I feel very good about our progress so far. I think we are carving a coherent plan out of the work to be done. And at this point we'd like to move to measures 249 and 250, outcome measures relating to hemodialysis adequacy, and Alan has reviewed both of these.

So, Karen?

DR. PACE: Yes. I just want to

bring this up and then we can move on. But we don't have any other new measures so the thought was to go back to our scheduled dialysis adequacy.

And I especially wanted to discuss 249 and 250 because they're basically the same measure with distinction that the last ESRD Committee wanted with the residual renal function. CMS has not been able to implement that, so they're bringing both measures back. And I think it's worth a discussion whether evidence has changed any that we need that measure specified that way or -- so, that's why I would like to have some discussion while you're all here about those two measures.

We can then decide if we want to continue on with all of the outcome measures in that group or if -- I'd like to just ask now if there are any other measures on our list that anyone has identified as a priority in terms of benefitting from discussion among the group?

If that's an okay plan, then we'll move on with dialysis adequacy. And we need to start with the measure developer intros to those topics.

CO-CHAIR CROOKS: Yes. Thank you. Thank you. Yes.

And our thought also was, perhaps, to try to get some vascular access discussion in this afternoon. Because we've done a lot of phosphate and mineral metabolism of late it feels like, so that may be where we head when we knock off some of the dialysis adequacy.

Lauren?

MS. RICHIE: Just one quick announcement. If anyone needs a shuttle this afternoon to the airport, BWI or Dulles, please see Tenee so that she can make arrangements with the hotel staff to have your shuttle arrangements for you.

CO-CHAIR CROOKS: Okay. Thank you.

MS. YERMILOV: Hi. I'm sorry to interrupt. This is Irina Yermilov, IMS

Health. And from what you just said all of our measures are under minimal metabolism. So can I assume that they probably won't be discussed today?

CO-CHAIR CROOKS: I'm sorry, what's your concern?

MS. YERMILOV: I am with IMS Health and all of our measures that were going to be discussed today were under mineral metabolism. And you just mentioned that you would probably go through dialysis and vascular access next. So can it be assumed that ours probably will not be discussed today under mineral metabolism?

DR. PACE: That's probably a safe bet. Could we email you if for some chance we think we'll get back to mineral metabolism?

MS. YERMILOV: Yes, of course. I don't know of Lauren is there. She definitely has my email address.

DR. PACE: Lauren?

MS. RICHIE: Yes. I'm here. I'll

email you.

DR. PACE: Okay.

MS. YERMILOV: Okay. All right. Great. Thank you very much.

DR. PACE: Thank you.

CO-CHAIR CROOKS: All right. So I'd like to invite CMS PCPI --

DR. PACE: PCPI.

CO-CHAIR CROOKS: -- PCPI, those two to introduce their candidate measures for dialysis for dialysis adequacy. CMS first.

DR. PACE: Yes, go ahead.

DR. MESSANA: It's my understanding we're talking specifically about 0249 and 0250.

DR. PACE: And we'll also --

CO-CHAIR CROOKS: The whole group.

DR. PACE: -- try to do the peritoneal outcome measures as well.

DR. MESSANA: Okay.

DR. PACE: So we'll try to focus on the outcome measures in this group.

DR. MESSANA: Okay. So very briefly because of time constraints and you want to get through a lot of stuff, I'm Joe Messana from University of Michigan, Kidney Epidemiology and Cost Center associated with Arbor Research as contract measure developers for CMS.

And the adequacy measures that we submitted were seven in total. Four related to hemodialysis adequacy and three related to peritoneal dialysis adequacy. But the centerpiece of all seven measures was the minimum targeted dose of dialysis for hemodialysis and peritoneal dialysis, respectively. Largely because those were the measures that are intermediate outcomes that are relatively proximate to a primary outcome. So they are the most important, and they contain the specifications from the corollary measures. So I think it's appropriate to focus primarily on those if short of time.

And the only other point that I

will make is particularly for hemodialysis but for PD as well, these types of measures have been reported for a number of years. And if you look at the CPM data there has been a progressive increase in the fraction of patients in the U.S. who have achieved these targets. And so one might be concerned that the performance gap criterion might be an issue. But we should keep in mind that most of the reporting of a very, very high fraction of patients relates to a subset of patients that have multiple values. So, it's a fairly constrained subset of people that have, for example, four values in a year in a facility. And so it may overstate the actual achievement. Some of the data that we included from CROWNWeb has a somewhat lower fraction of patients achieving these targets. So we believe there still may be a performance gap depending upon what data source you use and how you define the set.

And certainly because we believe

that this intermediate outcome measure is proximate to a primary outcome, we believe there is real risk of backsliding or regression if we do not continue to monitor closely this one of many, but one certainly measure of dialysis adequacy: small solute removal.

Thank you.

CO-CHAIR CROOKS: Thank you PCPI

MS. JOSEPH: Hi. I'm Diedra Joseph, again with AMA PCPI. Thank you again for the opportunity.

Our two measures are 0323 Hemodialysis Adequacy: Solute and 0321 Peritoneal Dialysis Adequacy: Solute. Both were previously endorsed by NQF and are being submitted for maintenance. And the most significant change to the measures, as you will notice, is the removal of the process component of the measure, which is the plan of care. The Work Group decided to focus on the intermediate clinical outcome for these

measures. And we have partially harmonized with the existing CMS measures. And our measures are specified at the physician level.

The measures have also been tested for reliability and face validity.

Thank you.

CO-CHAIR CROOKS: Thank you.

Okay. At this point I'd like to ask Dr. Kliger to -- I don't know if it works best to kind of put these up side-by-side or do you want to do them one at a time?

DR. KLIGER: We're going to set a record for accomplishment and time. So here it is.

Measure 0249, which is currently in place and we're being asked to renew it, is a measure of adequacy defined as all adults who have been on hemodialysis for six months or more and dialyzing three times a week whose single-pool Kt/V is more than or equal to 1.2 in the last measurement of the month using the Daugirdas or UKM measurements. This is what's

already in place right now.

Measure 0250, if I may I'll bring them up together, is the same measure but with the difference being that it excludes people that have greater than or equal to 2 milliliters per minute of endogenous renal function and it cuts it back down to three months instead of six months after starting dialysis.

The reasons that the second were introduced would seem pretty clear. The endogenous renal function is already incorporated, for example, in our PD measures. And that level of endogenous renal function is approximately equal to what three times a week 1.2 Kt/V would provide. So, it sort of would be a threshold.

The problem is that it's a completely untested measure. Even though it's there, we don't have any data on testing of that measure. And so I'll get back to that after we talk about 0249, but just so everyone

understands as we set the stage: With all the potential wisdom and the possibility of making it similar to what we do with PD, it's a measure that's untested and currently we're really being asked and required by the new, as I understand it, by the new standards of the NQF to examine the testing of a measure. So I suspect, at least I for one think we haven't the fulfilled the basic requirement to examine that one yet. But we'll get back to that.

So here in 0249 the single-pool Kt/V of 1.2. I want to just spend a moment looking, setting the stage for this.

Many of people have asked whether or not Kt/V urea is really is really the best test of adequacy, and that's really one of the underlying questions we have to address here. And if you're Dr. Ed Lowrie, you've been screaming for a while that it's the wrong measure. If you're Dr. Frank Gotch or John Daugirdas, you've been screaming for a while that there's no better measure and until a

better measure comes along, this is what we need to stick with.

Since this measure was first proposed and accepted in 2007 there have been really no substantial additional studies that would give us information on the question of whether this is the best measure or not, or anything more about that. So when we talk about the characteristics of the evidence, we'll really be talking, we'll be repeating the same discussion that was had in 2007. What's different now is that we have some testing that's been done that we'll have an opportunity to examine. So, that's the perspective, okay?

So why don't we go and talk about impact.

CO-CHAIR CROOKS: Okay.

DR. PACE: So it looks like the preliminary reviewers agreed it was --

DR. KLIGER: Sorry. Yes. Preliminary reviewers there say that the

impact is mostly high, and one person says moderate.

DR. PACE: Any discussion?

CO-CHAIR CROOKS: Anyone else? Okay. Let's vote for 1A impact; high, moderate, low or insufficient.

MS. RICHIE: Lorien, impact?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: Vote early and often. Okay. That's good. All right. Nineteen high, one moderate.

Next performance gap.

DR. KLIGER: All right. So as we just heard, that the developer quoted CROWNWeb, which is data from January of 2010 that examined this indicated that 66 percent of facilities -- and this is a facility level measure, incidentally. Sixty-six percent of facilities had 70 percent or more of their patients with that dose suggesting that, obviously, a third of facilities have less patients than that 70 percent who fulfill the

requirements.

CO-CHAIR CROOKS: And variability, is that also addressed in their submission? In other words, there may be some upper limit. Maybe 80 percent is the most you could ever do?

DR. KLIGER: Yes, I don't know the answer to that. I know somebody else may who looked at the data. But I'm just thinking of what Joe Messana told us before about their own data and the different ways of looking at it.

My interpretation looking at that is despite the fact that there's clearly been improvement, that there's still a performance gap.

DR. BERNS: One of the questions that I had that I've raised before is whether we should be using, or whether this measure should be at a single month value as opposed to several months. I don't think a rolling average is the right thing to do.

But thinking about practice and wanting to identify units or physicians that are outside of our expectations or outside of what would be considered quality, having a patient one month with a Kt/V below 1.2 doesn't tell me very much. Having a patient who is three consecutive months below 1.2 tells me a lot more. I don't know whether that's addressed in here, but whether that's something that we should be thinking about in trying to make sure that the measure does the right thing.

DR. KLIGER: Well, when we get to the specifications maybe we can examine that again.

DR. PACE: They didn't put it in 1B.2 about the distribution of performance, but I think on -- let me see if there was another place that they present it by quintiles. 2B.2.3.

DR. KLIGER: Right. Yes.

DR. PACE: There's information

about quintiles of performance that Lauren will bring up that will address your question.

DR. KLIGER: Yes. Peter was asking that, and it's there.

DR. PACE: Right.

DR. KLIGER: Another thing, when they look by quintiles it looks pretty tight. There clearly has been improvement. But there are the gaps. You see it right there on the screen.

CO-CHAIR CROOKS: Okay. Just as a question of process, Karen, are we being asked to pass one or the other or neither, or we could pass both of these that are so similar?

DR. PACE: Well, let me just give you the context, and I think Alan raised a good point about the next one not being tested and we are in a different place than we were back in 2007 where a lot of the measures were untested.

So, we want you to give us advice on this. I mean, if this measure for example

passes and it's adequate, and you agree that the other one's untested, it doesn't necessarily have to be recommended. That could be a recommendation for the next round that if that's really an improvement of the measure, that the next time the measure comes back that it actually captured the residual renal function.

So, I think we have multiple options.

CO-CHAIR CROOKS: Okay. And should they both pass, I guess then they'd be up as competing metrics and we could --

DR. PACE: Well, the way they had done it before, the way they were endorsed before is 0249 was supposed to sunset when they implemented 0250.

CO-CHAIR CROOKS: Yes.

DR. PACE: The problem is is that CROWNWeb never got going in order to implement 0250.

CO-CHAIR CROOKS: Okay. Thank you.

Stephen?

MR. McMURRAY: Peter, the other difference was the six months and three months time frame that's in there. And I guess I don't know whether you can have that discussion or not in here, but six months seems awfully long to start measuring this. And I have no reason -- I have no idea why it's that long, at least in today's current world. And so I don't know where that fits in the discussion of those two metrics.

CO-CHAIR CROOKS: Probably specifications would be the time to discuss that.

MR. McMURRAY: Right.

CO-CHAIR CROOKS: Okay. Thank you for clarifying that.

So we're getting to the point of voting on performance gap. Other discussion?

Alan, your light is on, does that mean you want the floor? Okay.

All right. Let's vote on

performance gap.

MS. RICHIE: And, Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: One vote for high, 19 for moderate, one low.

So we can go to the body of evidence.

DR. PACE: Yes.

DR. KLIGER: Right. Again, and the body of evidence is the same as the body of evidence was when this was first passed in 2007. It includes 11 or more studies that are retrospective observational trials showing a clear correlation between the dose of dialysis and heart outcomes, including in particular mortality.

There are no randomized prospective control trials looking at this, other than hemo. And all of you know that in hemo the test was between essentially what this current recommendation is and a modestly higher, a 16 percent higher dose. In that RCT there was no

survival advantage.

But all of the observational trials, as I say, have been clear and supported the fact that there's a correlation between outcomes, particularly in survival, and the dose. And that in many of the earlier trials 1.2 as a single-pool measure was picked because it was clear that at lower levels, and particularly at equilibrated Kt/Vs of less than about one, that the mortality was substantially higher. So the quantity of those studies, as I say, is over ten. And the quality, which we can go on and people can talk about this, are all really in observational retrospective trials.

CO-CHAIR CROOKS: Okay. So can we vote first on the quantity of studies in the body of evidence? High, moderate, low, insufficient based on our chart there. Go ahead.

MS. RICHIE: And, Lorien?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: Some of you may have voted too soon. There we go. We have 17 voting high and four moderate.

Okay. Now to the quality.

DR. KLIGER: Just one other thing that I will mention is that the DOPPS data, in particularly, if you examine it is not actually an RCT as you suggested before. But is very well done prospective work by facility and with stratification that makes it, I believe, very high level evidence although it's not an RCT. And that also has shown the correlation.

DR. BERNS: Alan, in all of these retrospective studies where is the breakpoint? My recollection is that it was really at one or 1.1.

DR. KLIGER: Yes. It's at one for equilibrated Kt/Vs. Single-pool Kt/V is about .2 higher. So a 1.2 single-pool is about equivalent to what the breakpoint is in the equilibrated.

CO-CHAIR CROOKS: Okay. Other discussion about the quality of the body of evidence? Jerry?

DR. JACKSON: A question for Alan. Does the DOPPS data duration of dialysis of a separate correlate with -- inverse correlate with mortality come into play or affect this measure at all or a totally a separate issue?

DR. KLIGER: Yes. With the DOPPS guy sitting in the back, I'm very reluctant. May I ask the developer to help us answer that question?

CO-CHAIR CROOKS: Sure.

DR. MESSANA: So there is a published analysis with Rajiv Saran first author from the DOPPS data that looks at duration of session after adjusting for Kt/V. And I can't remember if it was a equilibrated or single-pool Kt/V. Single-pool, Alan is telling me, which did show an independent effect of duration of dialysis session, and that's one of three or four observational

studies that show an independent effect of time after adjustment for single-pool Kt/V.

So the answer is time or duration of dialysis may be a separate predictor. But in my read of the literature it doesn't invalidate small solute removal as well.

CO-CHAIR CROOKS: Okay. Other questions, issues?

All right. Let's vote on the quality of the body of evidence; high, moderate, low, insufficient.

MS. RICHIE: And, Lorien, quality?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Okay. That's 21. The votes were six for high, 15 for moderate.

And on to consistency. Any discussion before we vote? Okay. Let's vote.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Ten voted high, 11 moderate. So this would pass the --

DR. PACE: Pass the evidence.

CO-CHAIR CROOKS: Pass the evidence.

DR. PACE: And it would pass importance.

Go to the next slide.

CO-CHAIR CROOKS: And it would pass importance, right. Do we need to vote?

DR. PACE: No.

CO-CHAIR CROOKS: No? Okay. All right.

DR. PACE: And we don't need to talk about that, okay?

CO-CHAIR CROOKS: So scientific acceptability.

DR. KLIGER: I have two comments and then I would really invite the others to join.

First, in terms of specifications. One point that we discussed at our last meeting was that this is a single-pool Kt/V rather than a standard Kt/V. And remember, the reason for that is single-pool is useful

if we're only comparing the same frequency of dialysis.

For all people on three times a week hemodialysis, it's reasonable to use this measure. However, we have an increasing number, although still relatively small but of patients going home, going four times a week, going five times a week, going six times a week. And at some point, and the developers do point this out, it would be useful to change from a single-pool Kt/V to a standard weekly Kt/V that will allow us to compare all of those different kinds instead of excluding people.

So in terms of the specs, my recommendation is that this is fine as it stands, but let's recognize that and let's urge developers as we move forward to look at measures that will help with different frequencies like the standard Kt/V. So that's one specification issue.

Then, Jeff, you had another one

about the frequency.

DR. BERNS: Yes. This is the same issue that I raised before with some of these measures where the patient variability or what have you, a single one month out of compliant to metric doesn't to me necessarily indicate that there's a quality problem. The identification ought to be, I think, around the people who are persistently below some value. If the Kt/V is 1.1 and you repeat and it's 1.4 or you -- that prompts a fistulagram and repair, then all the right things have happened. It's sort of what was talked about regarding the vitamin D and calcimimetic: If you respond appropriately, than that should somehow be a part of the metric, I think a performance measure.

CO-CHAIR CROOKS: Go ahead, Alan.

DR. KLIGER: I just want to move on with the reliability questions, because those are the specification questions. Are there any other --

CO-CHAIR CROOKS: Well, wait.

DR. KLIGER: Yes?

CO-CHAIR CROOKS: The three months versus six months versus one month, can we kind of clarify this for some of us how that all fits into the specifications? This is a monthly calculation, right?

DR. PACE: Yes. Right.

DR. KLIGER: Yes. I mean the rationale --

CO-CHAIR CROOKS: You want this to average it over three months or six months?

DR. KLIGER: No, no, no.

CO-CHAIR CROOKS: I'm not --

DR. KLIGER: I mean, the rationale originally was that you wanted patients to be stabilized and have appropriate vascular access and then have a reasonable measurement instead of doing it as soon as they start. But six months is a long tail and with the next measure, which hasn't been tested, it was suggested to reduce that down to three months

rather than to six. And, indeed, I think that's a good recommendation if we were to pass this one to ask the developers to consider making it three months instead of six months for this particular measure.

CO-CHAIR CROOKS: But as written it says six months?

DR. KLIGER: Correct.

CO-CHAIR CROOKS: Okay. Jerry?

DR. JACKSON: In addition to the type of vascular access and the duration after starting dialysis there's going to be facility variation according to how high turnover that clinic is. With a lot of new patients coming in, there tends to be a higher percentage of catheters in the early time frame, and that's going to slightly skew the results downward, where the facility that has a very stable population without turnover should be able to overcome that.

CO-CHAIR CROOKS: However, that sort of favors -- if you have a three month

window, that would be another factor kind of urging, addressing catheters early and often.

DR. JACKSON: Right. That may have been behind the idea of the six month. I don't know that.

CO-CHAIR CROOKS: Andrew?

DR. FENVES: Having said that, I agree with that completely. And, with fistulas failing at a higher rate than we thought of, at least in some studies suggest, that would put a disadvantage again if you had a lot of new patients because fistula -- another fistula, now you're outside the three month window easily.

CO-CHAIR CROOKS: Okay. But I personally feel either of those would negate shortening that window, in fact would put more attention on getting good access going at an earlier point.

Stephen, did you have any concerns? Okay.

DR. KLIGER: All right.

Reliability in testing in this case was done, as we've discussed before, by comparing two different time periods and showing a high piercing correlation of between .89 and .98. So the correlation is real good, but it's not quite the same patients and it's not quite the same time frame; it's somewhere in between.

So, it's tough but I guess my own thinking was I couldn't think of a better way to do this than that. And unless someone else had a thought about that, my sense was that in this case that's not a bad reliability test.

DR. PACE: Actually, and Lorien, if you could bring the measure developer responses. CMS did do some reliability of the precision of the measure score that they submitted back to us in response to questions. So, if you could bring that up on page 25. And what measure was this 0249.

Arbor, I was looking at this table and there's a measure number 0250, but was that really 49?

MR. PEARSON: Yes, that's correct. We apologize. Page 25 of our document.

DR. PACE: So page 25.

Do you want to just describe this?

DR. WOLFE: So we calculated some standard statistics related to signal-to-noise. And for 249, which is the one being discussed right now, the intraclass correlation was .34.

DR. PACE: Right. So in the table it's labeled 0250, but this is 0240.

DR. WOLFE: And we're sorry for that error.

DR. PACE: That's okay. I just wanted to get everybody on the right --

DR. WOLFE: And there are various statistics that are useful for looking at this. The r squared is .35 and this represents a highly substantial ability to distinguish between facilities. There are very substantial differences in a statistical sense, and you have also seen the distribution

of values across facilities with regard to their achievement of this measure.

So, both with regard to interclass correlation, which is good at .34, and the ability to see a signal between facilities in the face of patient-to-patient variation this measure is very successful.

DR. KLIGER: Okay. Again, just to wrap this up from my perspective unless there's anyone else that had comments, the reliability asked us about the precision of the specifications. I think they're precise. We might have suggestions for altering them, but they're precise and you've just heard the rest of the reliability.

CO-CHAIR CROOKS: So let's vote on specification and reliability. Is everyone ready? Okay.

MS. RICHIE: Lorien, reliability? Lorien?

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

MS. RICHIE: Okay.

CO-CHAIR CROOKS: Okay. That's 21. Twelve voted high, nine moderate.

Validity?

DR. KLIGER: Actually, the validity was looking at the quintiles of performance compared to SMRs. And here's where I'm going to invite Janet Welch to make some comments, because she was the one who had the most concerns about this. But overall if you look at the numbers, what it appears to be is that compared to the highest or that is the best quintile, all of the others had statistically significantly worse mortality. It was not really well graded, it wasn't like the very worst mortality was the lowest quintile and it graded up from there. But clearly the four less than optimal of the quintiles had a higher mortality than the highest quintile. So those were the validity data that were presented.

Janet, do you want to say some

words about that?

DR. WELCH: That data looks like it's curvilinear and I couldn't make sense of that in terms of validity data.

DR. FISCHER: But it's just maybe there's a nonlinear relationship. I mean, I just may be that there's a nonlinear. I also tend to think linearly, but there are a lot of nonlinear biologic relationships.

DR. KLIGER: Right. But I must say, again, when we first talked about this measure and when it was first developed we had no link, really, no effective link in testing between the measure and hard outcome like mortality. This actually provides some of that data that is very helpful to -- at least to me.

DR. FISCHER: Once again, this is kind of one of these indirect measures of validity, right? I mean, in other words, the face validity is this really measuring what it's supposed to be measuring remains

unanswered. I'm not saying I have a better idea; I don't. But once again this has kind of come up, Karen, a couple of times. And it seems like overall the Committee this has been sufficient.

DR. PACE: Right. I mean, validity is not definitive by any one test. It's something you kind of build on over time. And when you're talking about especially the measure score, I mean what we're most interested in is if you have a group of providers and you have scores, can you say this provider is better than that one because they have a better score than that one. We really want to be able to make valid conclusions about quality. And they're saying that one way that you could do that, because outcomes are what matter, people dying or living and showing a correlation between having a score on this measure to score on the mortality rate, it provides some demonstration that you're going to be making some valid

conclusions.

I invite others to kind of add to that discussion. Jerry?

DR. JACKSON: I almost hate to bring this up, but we struggle with it at the networks. It's fairly well know that dialysis staff will encourage patients to stay on their full fully prescribed time the one day of the month this is measured and often throughout the month patients sign off early. So the only way to overcome this would be to get an average single treatment Kt/V, which is really not very feasible, I don't think. So this is probably the best we can do. But I think that that might --

DR. KLIGER: It's a nonlinear function. You can't get an average. Kt/V will not be a valid measure, really.

DR. JACKSON: And that might explain some of this nonlinear in the quintile to support that.

DR. PACE: But I think that speaks

to the issue of validity. You know, because it is I think the last measure of the month. And so, you know that is definitely --

DR. JACKSON: Well, at least done once a month.

CO-CHAIR CROOKS: I think that's a very interesting observation. I guess you have to just kind of hope that the game playing goes on about the same frequency at all units, you know. Because I don't know how to get that out of there.

DR. JACKSON: I think it's signal-to-noise, really.

DR. KLIGER: Well we could be like CMS and walk in there and do a surprise visit and measure it unexpectedly. But, that's not going to happen.

CO-CHAIR CROOKS: So I think that's a threat to validity, but it's one that we can't eliminate, and I don't think it overrides. Does it override the value of the metric?

Okay. Other thoughts or issues before we vote on validity?

DR. PACE: So let me just point up here. I guess the question about you'll be voting on the measure as specified, which is the per month. And so the issue about wanting to change the metric, and it sounds like that's a validity question for you, Jeff, about doing a single measurement versus persistent. So your vote on this if it passes here would make the measure go forward as it is. So I'm just going to point that out so that we know.

I know that we had that discussion on several metrics in the last project. Ultimately they ended up going through as they were originally specified.

DR. KLIGER: I'm sorry. Unless I missed it, the measure doesn't specify how frequently it should be measured. It doesn't say a month.

DR. PACE: No, but isn't it a

single measure per month?

DR. KLIGER: So it can be -- it gives a numerator and a denominator and it says in the study period. Unless I've missed it, it doesn't say. It can be three times a month, it can be -- you know, it's whenever it is measured, this is the way to do it.

DR. PACE: Okay. So, Bob, you want to clarify? Because Jeff's point was he was bringing up the persistent over several months, right? Okay.

So is it one measure per month, Bob?

DR. WOLFE: A couple of issues.

It is specified -- I believe it is specified and it's intended to be specified as just one measure per month, and it would be the last dialysis session of the month.

DR. MESSANA: In 2A.1.1 numerator statement, I think it says here, the parenthetical statement "Is calculated from the last measurements of the month using urea

kinetic."

DR. WOLFE: So there's another question of what is the duration of the study period. It is intended so that it could be meaningful just with one cross section of one month measured at anytime of the year. It's expected that it may be reported for longer durations as well. But it is proposed that each patient month count equally as one patient month.

So a patient who was there for six months would contribute six patient months and a patient who was only being treated at the facility for one month during that study period, would contribute one patient month.

I think this is very similar to the discussion that took place yesterday that if they're out of alignment for just one month, that would have less of an impact over a six month study period than if they were out of alignment for all six months.

DR. KLIGER: Okay. I would suggest

that we look very critically at the way it's actually written. Because in my view it does not say it's done every month. And if that's the intention, we just should make it clear that that's it.

DR. PACE: Right. And it also is specified where it looks like patient is the unit versus month as the unit, as you were just describing.

DR. BERNS: It would be very helpful to me, maybe, if you could do the analysis in a way that if you look at, for instance, SMR and Kt/V below 1.2 for three consecutive months. An whether that is a better predictor of mortality. Because the hazard ratios, if that's what it was, were statistically significant but small because of the large number of patients that you measure.

So, if there was much better discrimination by tweaking the measure a little bit, I think it would be more useful to us.

DR. WOLFE: Thank you very much for the suggestions.

And we have looked at that for some of the other measures, whether to roll them up and say are they persistently low with regard to the outcome and in particular for anemia. But I don't know if we have same analysis for Kt/V. I believe not. But that is something to investigate. Thank you.

CO-CHAIR CROOKS: I might comment, though, because this is a facility level metric, you might catch a patient here and there on a bad month or there may be some variability, but you would think that might average out in the statistics.

Okay.

DR. PACE: Question.

CO-CHAIR CROOKS: That got raised hands. Yes, Bob?

DR. WOLFE: One more clarification for Alan. The current implementation that is planned to my understanding is month-by-month.

So every single month there would be a report, which would have one month's worth of data in it. So that's --

DR. KLIGER: No, no. I get that and, in fact, of course that's what we've all been doing for many years. I'm just saying that when I look at this specification it's not so clear here.

DR. WOLFE: Thank you.

CO-CHAIR CROOKS: Okay. Are we ready to vote on validity now? Any other questions? Okay. Let's vote.

MS. RICHIE: Lorien, validity?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: That's 21. We have two voting high and 19 moderate. Okay. So I think we passed the scientific acceptable of measure properties. Do we need to look a disparities in this case?

DR. PACE: Yes. I know we've been kind of hit and miss here, so I apologize. And I don't remember if we discussed it under

performance gap if there were any disparity issues.

DR. KLIGER: Right. There were none that were described and they show us the performance in various strata with no evidence of disparities.

DR. PACE: And it seems like because CMS is using race data for the mortality measure, they have the data that could be applied here if needed to look at differences by race. Okay.

CO-CHAIR CROOKS: Yes. The analysis can certainly be done, but there's no reason to think that --

DR. PACE: It has to be.

CO-CHAIR CROOKS: -- race impacts the dialysis prescription per se.

DR. KLIGER: So I would suggest this question is not relevant.

DR. PACE: Okay.

DR. KLIGER: Because it's an "if" question.

DR. PACE: Right. Right. Good.

CO-CHAIR CROOKS: Okay. Onto useability, Alan.

DR. KLIGER: So just really quickly, it's been in use for many years and the evidence is that it is useful.

CO-CHAIR CROOKS: Thank you for being succinct.

Any other comments on useability either for public reporting or quality improvement? I think we're ready to vote then; high, moderate, low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: So we have 17 voting high, four moderate.

So we can move to feasibility.

DR. KLIGER: It has proven to be feasible.

CO-CHAIR CROOKS: Ah, that was two words less than the last time. Okay. I think that's a pretty solid rationale. Others?

Okay. So let's vote on feasibility. High, moderate or low, insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: We're stuck on 20. Oh, there's 21. Okay. So we have 21 high. Very feasible, apparently.

Okay. So the next one is the overall, and we do need to vote does the measure meet all the criteria to be suitable for endorsement and to review. I think each section it has passed. So yes, no or abstain. Let's vote.

MS. RICHIE: And, Lorien?

DR. DALRYMPLE: Yes.

CO-CHAIR CROOKS: So we have 21 yes.

So, to 250. We need to go through the same --

DR. KLIGER: So if I may, my recommendation is this: I don't think that

this measure has the characteristics that will allow us to vote on it because it's been untested.

DR. PACE: And I think that's an excellent observation. And so it could not pass reliability and validity, so there's really not much point, other then I guess whether we want to make the recommendation -- I'd like to at least have a discussion about whether it's valuable to add the renal residual function into a measure for maybe the next iteration.

DR. KLIGER: Yes. So maybe I can start that discussion, and it's a good discussion. Because it would make logical sense to do that. However, what's interesting is that I haven't seen any data that suggests that with or without factored in effects any measured outcomes or change in outcomes. So if there is such evidence, it would be useful for the developer to bring that to us.

CO-CHAIR CROOKS: Isn't there

evidence that -- well, I guess that's really not relevant to the point.

DR. KLIGER: Endogenous renal function is good. No question about that.

CO-CHAIR CROOKS: Yes.

DR. KLIGER: But the question is whether this particular measure of adequacy is better if you factor in endogenous kidney function or not. My gut says it should be, but I'd like to see some evidence.

DR. PACE: If I recall from the last project, the Committee had suggested that be included along with shortening the time frame. I guess that was one of their issues of shortening the time frame you might be capturing patients that still had -- so I don't know. But I think that's a good question, an outstanding question whether it really improves the measure.

CO-CHAIR CROOKS: Well, I think from earlier discussion I think the sense of the Steering Committee was three months was

better. And we'd all like to see some data about the usefulness of putting that into the metric.

DR. PACE: Right.

CO-CHAIR CROOKS: Putting the residual renal function into the metric.

DR. PACE: So, I guess let us go before we resolve that question, the current measure that we just passed, 249, is specified with after six months, right? And are you recommending that that be changed to three months, and is your recommendation continued on that point? Bob or Joe?

DR. MESSANA: Just one comment to reenforce the data that was presented and was discussed was for the six month exclusion measure. That's what you've reviewed today, to this point.

CO-CHAIR CROOKS: So we're not recommending that they consider changing that particular --

DR. KLIGER: No, no. I wouldn't say

that. I mean, Joe is of course exactly right, so we passed the right measure and we looked at the data for the right measure. But listening to what my colleagues on my right here said earlier, I do think it would be wise to ask them to consider if there is evidence to moving that to three months.

CO-CHAIR CROOKS: Would we have to look at more reliability data or anything for them to do that, or could they just make that change and still be an endorsed metric?

DR. PACE: I guess that would be a question for you all. What would be the downside of having a shorter -- I mean, we talked about the upside that it's getting more patients in there, it provides an incentive to get the vascular access, but what's the potential downside?

DR. BERNS: If I understand correctly, then the relationship between Kt/V and SMR was based on the six month time frame.

DR. PACE: Right.

DR. BERNS: So we would need to see that the same relationship held with the same statistical significance and so forth at three months. And until we see that, I think it's hard to make a decision that a change should be made.

DR. PACE: Okay.

CO-CHAIR CROOKS: So if I'm catching your drift, than we probably should not encourage them to change it because we'd have to look at some testing of the data?

DR. BERNS: Well, I would encourage them to look at that data.

DR. KLIGER: Yes, that's right. I agree.

CO-CHAIR CROOKS: Say that again.

DR. BERNS: I would encourage them to do the analysis of three months with SMR or some other outcome and then come back and it may be a stronger relationship for all we know.

CO-CHAIR CROOKS: So if that could

be accomplished in the next month or two while we're still in operation?

DR. PACE: Right. So I guess where we would stand is the measure as it is can move forward, but we're going to put in a request to CMS and their contractor if they could do some analysis of changing that time period to three months and we would be especially interested in looking at that relationship to SMR? Would that do it? Okay.

CO-CHAIR CROOKS: Okay. All right with everybody?

DR. PACE: And unless anyone objects, we will not go any further with 250 because that measure is not tested. And if turns out that that's a better way to do it when they bring the measures back for the next round of maintenance, they should incorporate that. Okay?

CO-CHAIR CROOKS: Okay. Thank you, Alan, for guiding us through all that.

And I think we'd like to go to 323

next, the PCPI metric on Hemodialysis Adequacy: Solute, which is a --

DR. PACE: Physician.

CO-CHAIR CROOKS: -- reendorsement?

DR. PACE: Yes. And it's also a physician level.

CO-CHAIR CROOKS: A physician. And this was assigned to Michael Somers.

MR. SOMERS: So this is a measure up for renewal. It's looking at dialysis patients in the percentage of calendar months within a 12 month period when they have a single-pool Kt/V greater than or equal 1.2.

I think a lot of the general discussion that we just had on the last measure is going to be very applicable to this as well.

If we look at impact, four of the five reviewers assigned it high. The measure stewards also included some newer citations with evidence since the initial endorsement to reenforce the impact.

DR. PACE: Okay. Shall we vote on impact and then we can move on to discussing the rest of the measure?

MS. RICHIE: And, Lorien, impact?

DR. DALRYMPLE: High.

DR. PACE: Okay. Anybody? Okay. Go ahead.

CO-CHAIR CROOKS: Twenty voted high, there were no other votes.

MR. SOMERS: Okay. In terms of opportunity improvement, although the measure developers acknowledged that the percentage of patients achieving this has been increasing, they did give evidence of a performance gap, not only between men and women, they also quoted some older racial data that had been in their initial application as well. I think that data is still probably applicable even though there was newer data mentioned in this application.

They also alluded to some CMS PQRS data showing that 41 percent of patients

didn't meet this standard in the period that they reviewed.

CO-CHAIR CROOKS: Thanks. Any other comments from the reviewers or the Committee? Okay. Let's vote on the performance gap.

MS. RICHIE: Lorien, performance?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Okay. The voting: Four high, 17 moderate. So this is not an outcome per se?

DR. PACE: Right.

CO-CHAIR CROOKS: So we go to the body of evidence then.

MR. SOMERS: So in terms of quantity of the data they go back to the KDOQI guidelines. They allude to 87 articles that were abstracted or that were used initially for that guideline and 23 studies that were then used for the summary tables within that guideline.

They also had some more specific

comments about the hemo study along the lines of what Alan discussed with the last measure as well.

CO-CHAIR CROOKS: First we'll vote on quantity of studies. Any other discussion.

Okay. Let's vote.

MS. RICHIE: Lorien?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: That's 21. Seventeen voted high, four moderate.

Next is the quality.

MR. SOMERS: Again, I think our discussion with the last measure would be germane here as well. There was only the hemo study that was a minimized control study.

CO-CHAIR CROOKS: Okay. Any other comments? Okay. Let's vote on the quality?

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: That's 21. We have five votes for high and 16 for moderate.

And now the consistency.

MR. SOMERS: Again, I think our comments from the last measure would also be applicable here since it's the exact same data.

CO-CHAIR CROOKS: Good. All right. Let's vote on consistency.

MS. RICHIE: Lorien, consistency?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Okay. Four votes for high, 16 moderate and one low.

So this would pass with a medium, moderate or high level for all three.

DR. PACE: Yes.

CO-CHAIR CROOKS: So it does pass the evidence decision logic grid.

DR. PACE: Right.

CO-CHAIR CROOKS: And so the next question is does this pass the importance. And because it did pass all three --

DR. PACE: Yes. Tenee, will you change it? Okay.

CO-CHAIR CROOKS: And --

DR. PACE: Right. You have to go back to the importance. Yes. So it passed it all three.

CO-CHAIR CROOKS: All three were met. So I don't think we need to vote.

DR. PACE: No.

CO-CHAIR CROOKS: Unless the Committee feels differently. Okay.

So let's go on to scientific --

DR. PACE: Reliability.

CO-CHAIR CROOKS: -- acceptability, reliability and specifications.

MR. SOMERS: So the reliability was tested by some data extractions from patient records from four clinical sites per the PCPI Testing Project. And they showed a reliability that was 99.7 percent. It was inter-rater reliability that they were essentially testing.

CO-CHAIR CROOKS: Yes, that's a good reliability test I think we would say, right?

DR. PACE: Yes. As we discussed yesterday, the main issue is that it was tested with inter-rater reliability in terms of extraction but it's been implemented with CPT II codes and they are proposing electronic record specification. So there's a little bit of a mismatch there.

CO-CHAIR CROOKS: Disconnect?

DR. PACE: But again, you'll have to apply your judgment to that.

CO-CHAIR CROOKS: Although one might think going from -- well claims data has its own issues.

DR. PACE: Right.

CO-CHAIR CROOKS: But going to electronic might also be an advantage.

Any specification concerns?

MR. SOMERS: Similar to some of the measures we discussed yesterday when you go into the PDF that came with the initial measure and some of the diagnosis included things pertaining acute dialysis and not a

chronic dialysis.

CO-CHAIR CROOKS: Okay. I don't recall, were electronic specifications submitted with this measure as well, and did anyone look at those?

MR. SOMERS: That was what --

DR. PACE: I'm sorry. And did you identify any issues with it?

MR. SOMERS: There were, again, like several of the measures yesterday.

DR. PACE: Okay.

MR. SOMERS: Codes that correlate to continuous forms of dialysis and more acute kidney injury settings for dialysis.

DR. PACE: Okay. So do we need to kind of separate those out for now and ask PCPI to come back -- okay. Thank you. I'm sorry.

CO-CHAIR CROOKS: Because Kt/V isn't usually measured on an acute patient. Do they sort of come out in the wash anyway? I guess I don't know. No one can tell us that

for sure. Okay.

So we'd like to have some review of the CPT code selections.

All right. Other issues? Assuming that's done, shall we vote on reliability and specifications?

DR. DALRYMPLE: This is Lorien. Can I ask one question on this proper reliability.

CO-CHAIR CROOKS: Yes. Yes.

DR. DALRYMPLE: And is this is using the CPT II codes on the performance of CPT II codes?

CO-CHAIR CROOKS: I couldn't understand you very well.

DR. DALRYMPLE: Oh, I'm sorry. I know for some of the other measures there was data available on how well the CPT II codes performed. And since they're proposing to implement this using CPT II codes, is there any data they could provide us on the reliability of the CPT II code as opposed to

the chart review, which does not appear to be the primary way that it will be implemented?

CO-CHAIR CROOKS: Can you answer her concern?

MS. CHRISTENSEN: I'll clarify again that the primary way it's going to be implemented is we are not recommending a primary way of CPT II codes. It is an option, just like the other measures.

We did provide some data in there somewhere on the reliability is over 50 percent for the comparison between CPT II codes and going back in and manually extracting. But, again, it's the same problem with billing on a monthly cycle and the billing cycle may not be on the same cycle as the actual calendar month. So it's really hard to say just because of the way the program's implemented.

DR. DALRYMPLE: So is the primary way that this is going to be recommended to be implemented by manual chart review or by EHR?

MS. CHRISTENSEN: PCPI does not make a recommendation as to implementation. We would simply provide the specifications for all available forms of implementation.

DR. PACE: Right. But the reality right now is this is being implemented using CPT II codes, correct? And is there any plan to implement it widespread using medical record abstraction?

MS. JOSEPH: We simply asked our specifications team to supply all of those different specifications for EHR, for paper and for claims. But we're not sure how people will choose to implement them. It is an option.

DR. JONES: I mean, it is good to defend your point of practice level. Where the practice level isn't that point, if there's still paper, they'll do paper or they can have electronic. But the specifications are meant to be able to let them do it electronic.

DR. PACE: Right. And that's on an individual practice choice. But when we're talking about endorsing measures, it's from the standpoint that they will be used for both public reporting and quality improvement. So it does make a difference for standardization standpoint.

If you were going to use this in your own practice for quality improvement, you could choose whatever works for you.

So, yes?

MS. CHRISTENSEN: I mean, I guess all I can say for that is CMS does run the PTRI/PTRS program, so that's not our actual program. But we have historically that they go from claims-based measures to registry and EHR-based measures. So I don't know their thinking personally, but that is certainly a possibility that they might choose to do that.

DR. PACE: Right. I mean in general the idea is for all of health care to move toward electronic record measures. So I

mean that's the push from CMS, HHS, NQF is very much involved in that. So, I mean that's the goal and we'd like that. But the current status is in terms of these programs, and I don't know you may know more than I do in terms of what the kind of projected time line is for CMS. And I have no idea about that.

DR. WELCH: I did have some questions about computation of the variable, because I am just looking at my note here. Is that the denominator in the text is that it's all calendar months that patients are receiving hemodialysis three times a week. But on the e-specification document the denominator is all patients identified with an initial patient population. So they don't seem like the values are the same. Did I miss something?

MS. CHRISTENSEN: I think we already divorced the e-specifications, right? But we definitely are interested in your feedback on those e-specifications.

DR. WELCH: Okay. All right. Oh, I missed that.

MS. CHRISTENSEN: It's tough to do them.

DR. PACE: So we're separating those out for now. We'll come back to it if they can with the crosswalk. Otherwise, for now we'll be considering the measure with the medical record at the CPT II code specifications.

CO-CHAIR CROOKS: Jeff?

DR. BERNS: The question that I had, the prior ones from CMS were facility level. This, if I understand it correctly, is position level. And I'm not sure that the reliability or validity has been tested at the physician level. In other words, it is actually the right physician that's attached to that specific Kt/V value?

DR. PACE: That's what they provided in their submission is testing at the physician level in four practices, I believe.

But you know you bring up just a point that we'll have to deal with on a harmonization issue. These measures are specified differently. And the question is, you know the facility level measure that we talked about is a patient level, this is months. So we'll have to have a discussion about that whether that presents any problems with interpretability, et cetera. But we'll set that side for a later discussion when we get to harmonization issues.

CO-CHAIR CROOKS: So my take away at this point of the discussion on reliability is that chart extraction method has been tested and found reliable. We're expecting that in the long run this should be done more in electronic data, which is a good thing and is generally reliable but hasn't been really tested fully. Is that a good summation?

DR. DALRYMPLE: Well, but what about the CPT II code finding if people actually chose to implement this using CPT II

codes instead of one of the options being proposed? And reliability does not seem very strong to me if I understand the data correctly. But I'd be interested in how other Steering Committee members interpret those statistics.

DR. PACE: Right. So this is they presented it under comparability of multiple data sources or methods and to be fixed, I think -- is that what you're referring to?

DR. DALRYMPLE: Well, I think if I understood the steward correctly when they looked at CPT codes there was slightly higher than 50 percent reliability because there continued to be issues of claim forms lagging monthly, if I understand correctly. One of the issues that came up with measures yesterday that it seems the CPT II codes have some limitations because of monthly lag and that there are some issues of reliability when you use them. But please correct me if I'm misunderstanding the presentation of the data.

CO-CHAIR CROOKS: Yes?

MS. CHRISTENSEN: If I may, we're not suggesting that the data reported by the practices using CPT II codes is in anyway wrong. We're just suggesting that because of their monthly billing cycles the way our abstractors looked at it and the way they were reporting it was different. But this month -- the month blocks --

CO-CHAIR CROOKS: It's a different month, right.

MS. CHRISTENSEN: The patient months were the same, we just were looking at different patient months then the patient months they were looking at if that helps.

DR. PACE: And I think the other point about this -- and that's where you had the 64.9 percent agreement? Okay.

The other thing to point out, this measure has changed from the time of endorsement. And so this testing was the prior measure that had the plan of care component,

which was problematic anyway. But I don't know, do you have any sense of how this would play out with the revised measure?

MS. CHRISTENSEN: Yes. One thing that I will say is that we do see more and more physicians using the measures every year. So they must be getting something out of them. I wish we could provide more data, but CMS is not able to provide it yet.

DR. BERNS: I hate to belabor the point, but I'm not seeing where it's documented on an individual physician level the reliability --

DR. KLIGER: So you're talking about attribution, really?

DR. BERNS: Yes. Yes, is it Jeff Berns seeing that patient that month or is it reliable at the facility level or the shift level? Maybe I'm just not getting something.

MS. CHRISTENSEN: To speak to the PQRI program, I believe that the physicians self-report for their own patients. So that

isn't a problem in the PQRI program, if that makes sense, the way the measure is done.

DR. JONES: The individual charts were done by physicians, went into that physician's chart, they extract the information to see if it was congruent. So it was done through that individual physician, not through the group. That's how the extraction happened with all the ones we presented.

DR. PACE: So with chart abstraction, obviously, you're not doing any kind of algorithms to see which patients belong to which physicians. You're going to the physician's office and looking at charts. With the PQRS or PQRI program, physicians are self-reporting. So that's all we know at this point.

CO-CHAIR CROOKS: Stephen?

MR. McMURRAY: Peter, in the practices around the country there is such variability of who sees a person in a dialysis

facility month to month, that I'm not certain going to the facility and looking for that month validates anything. Because the next month it may be someone else seeing that patient, or for three months. I mean, the practice variation is enormous around the country of how actually this all takes place. And so to rely on just that chart abstraction on a few practices seems to me to be -- I'm not sure how helpful it is.

CO-CHAIR CROOKS: Can we clarify? Is this at the physician level or the physician group level? Because would that take care of your concern if that was the case?

MR. McMURRAY: It would be better.

CO-CHAIR CROOKS: It would be better?

MR. McMURRAY: It would be better. It doesn't get you to a physician level, but there is a marked variation in physician practice patterns in the facilities.

CO-CHAIR CROOKS: Ruben?

DR. VELEZ: But if they picked PQRI, I understand that a patient is assigned to a physician and it would go under that physician which means, on the other hand, I may be seeing a 100 dialysis patients but they're not assigned to me. I would not be doing that. You know, so I'm not sure any measure will be able to adapt to the practices. There are 500 different ways of practicing in the U.S., but that's what I think is the PQRI process.

CO-CHAIR CROOKS: And also, if you're rounding on someone else's patients and you're not doing a good job, it's their job to put some pressure on you, hey, you're seeing this patient, you know, so they can feed back to you and say you'd better tweak their dialysis prescription.

Does that answer your concern, Stephen?

MR. McMURRAY: In very few

practices does that happen because, you know the discontinuity of what's going on isn't --

CO-CHAIR CROOKS: But this would make you, perhaps, put in a system to help monitor each others' behavior. Might be a good thing.

Okay.

DR. JONES: And again, for the measure, and I think this happened yesterday too, are we asking the reliability that what's happening out there in the field now, can this measure get out of the physician's chart in what they're trying to put in? So I'm not sure we're ever going to solve the problem you have here in the near future, but with the tools that we have now is this measure going to accomplish what we can do in today's world. And I think that's what the question is. And I think going through at least a chart abstraction, going into a physician's office, pulling out that information is about as good as you're going to get for the state of the

art today.

CO-CHAIR CROOKS: Okay. Are we ready to vote on reliability? Okay. Hearing no objection, let's vote.

MS. RICHIE: Lorien?

DR. DALRYMPLE: For reliability low.

CO-CHAIR CROOKS: That's 21. We have 17 voting moderate, 2 low and 2 insufficient evidence.

Okay. So both validity and reliability have passed --

DR. PACE: No, we haven't voted yet.

CO-CHAIR CROOKS: Oh, that was reliability. Let's move on to validity.

MR. SOMERS: So they used face validity, they had a panel of 19 experts, mean rating 4.63 over five.

DR. PACE: And this is where we would also ask if there are any exclusions for the measure, and if that had been -- any

analysis on exclusions.

MR. SOMERS: I didn't see any exclusions.

DR. PACE: So this has the general exclusions of the --

MR. SOMERS: Well, it did say in it somewhere about medical or system issue in a flow chart somewhere. It didn't say anything in the narrative.

DR. PACE: Right. And in the specifications, it says: an exclusion is some documentation of a medical reason for the patient not having achieved 1.2 or greater.

And let me see what -- if you'd go for the specs for exclusions. And the details just say that -- they give one example. Patient has residual kidney function. Then other medical reasons. And then from the CPT coding standpoint, they amend, they put in a modifier that says the patient had an exclusion. But they didn't have any analysis because they added that exclusion after they

had done the testing.

CO-CHAIR CROOKS: Okay. Is anybody concerned about that or like to discuss the validity? Okay.

So let's vote on 2B, validity; high, moderate, low or insufficient evidence.

DR. DALRYMPLE: I'm sorry. Before we start the voting, this is Lorien, I was disconnected. Did you already start the voting?

CO-CHAIR CROOKS: We're just voting now. Yes, we'll restart the voting. We were just voting on validity.

DR. DALRYMPLE: I just wondering if you would mind just giving a brief summary of the Committee's thoughts on validity? I apologize for getting disconnected.

CO-CHAIR CROOKS: Michael, will you give the high level?

MR. SOMERS: So we talked about the face validity being used for the measure. And we also talked about there being some

denominator exclusions for medical reasons, although the validity of that hasn't been tested.

DR. DALRYMPLE: Okay.

CO-CHAIR CROOKS: That was added after the testing, that exclusion.

Okay. So let's vote validity: high, moderate, low, insufficient evidence.

MS. RICHIE: And Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Okay. That's 21. We have 18 votes for moderate, one low and two insufficient.

So now I think I can safely say that we have passed the scientific acceptability of measure properties.

Disparities, back up one side. Again, this is similar to the last measure. We don't think there's reasons that there should be disparities and the data could be examined that way for disparities, right?

DR. PACE: I assume they didn't

identify any disparities or --

MR. SOMERS: Well, they did allude to the PQRI data with the 50 percentile, or 50 percent of physicians having performance between 30 and 80 percent.

DR. PACE: But no differences by race or --

MR. SOMERS: Just general allusions as to there being a performance gap by race.

DR. PACE: Okay. Any reason to vote on this on disparities?

DR. KLIGER: Yes. What I just heard was that there was a disparity by race.

DR. PACE: Okay. All right. And the measure is not --

DR. KLIGER: What was the disparity that you're describing, Mike?

MR. SOMERS: When they were talking about, back in the section about high impact and in opportunities for improvement they alluded to data from the `90s about how there was differences in achieving Kt/V goals in

African-Americans versus other populations.

DR. KLIGER: So 1894 or --

MR. SOMERS: No. I don't know. I think it was data from '93 and '97. It was in their original application and they didn't have any newer data in this.

DR. PACE: Okay. We'll move on.

CO-CHAIR CROOKS: Okay? All right. So to feasibility -- usability.

MR. SOMERS: I think like before it is used.

CO-CHAIR CROOKS: It is used.

Okay. So any other discussion about usability? Okay. Let's vote: High, moderate, low, insufficient.

MS. RICHIE: And, Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Fourteen voted high, seven moderate. So it passes usability.

Let's go to feasibility.

MR. SOMERS: It is feasible.

CO-CHAIR CROOKS: Could you shorten

that up a little bit? 'Tis feasible, maybe?

Okay. So this is being done, although it's going to change a little bit.

Any other discussion or comments? All right. Let's vote. Feasibility.

MS. RICHIE: Lorien? Lorien, feasibility?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: We have 13 voting for high and eight voting moderate.

So let's go to the next slide then. It has passed all four areas.

DR. PACE: Right.

CO-CHAIR CROOKS: So let's have the final vote. Does the measure meet all of the criteria to be suitable for endorsement; yes, no or abstain.

MS. RICHIE: Lorien?

DR. DALRYMPLE: No.

CO-CHAIR CROOKS: We'll wait until we're done with the votes.

DR. PACE: Everybody voting?

CO-CHAIR CROOKS: I guess that's going to be -- oh, there's 21. Okay. So 20 yes, one no.

Alan, you had a comment? No? Okay.

So that completes this metric.

We still have 20 minutes before the planned lunchtime. I wonder if we could -- should we go to 321?

DR. PACE: Let's go to public comment.

CO-CHAIR CROOKS: Oh, public comment.

DR. NALLY: Can I ask a quick question? And I didn't want to bring this up, but Alan started us out alluding to the controversy of ways to measure adequacy; URR, Kt/V. We have been used to Kt/V for a period of years now, but as recently as instituting the QIP, CMS had in the URR which seems, I guess, to be going away, was there science to that transition or just recognition of the

obvious?

DR. KLIGER: Thank you very much. You know, I think the more relevant question is whether urea kinetics is really the way to go altogether. Is time alone, is frequency alone, is volume alone a better predictor of outcomes than is urea kinetics? Those are really the hot issues that people are taking a really careful look at now.

When you look specifically within urea kinetic modeling, there are several ways to do that. And if you speak to the experts, they do tend to agree that URR is not the best measure and probably one of the more specific measures, UKM or Daugirdas or one of those is probably better.

DR. PACE: We'll do public comment, get lunch, we'll take a little break and try to resume a working lunch. And given the time frame, I think after lunch we'll move on to vascular access because we have some other measure developers here that --

CO-CHAIR CROOKS: Did you want to do this?

DR. PACE: No. I think we'll just, so that we get a little discussion about another topic area before we dispense with everyone.

So let's go to public comment. And first of all, is there anyone on the phone that wants to make public comment?

Okay. Peter, I'll let you --

CO-CHAIR CROOKS: Okay. Hands.

DR. JONES: On behalf of PCPI.

I would be remiss not to go back to yesterday's discussion, particularly with this being the last of CKDESRD review, I think in the next number of years, even though you mentioned yesterday there could be a period where things could be relooked at. But we're talking about potentially a couple of years before we do this. And yet we may leave this setting without having an important safety metric, and I'm talking about trying to

prevent -- or recognizing an increasing incidence of transfusions in patients with an anemia management. And without having a lower level, whatever that might be, to try to help all of us make sure that our patients are not transfused. And I'm concerned that we did not have -- and that would be obviously the fault of those of us who did not present the information, all of the information in front of you, particularly with some of the data. Although it not being well controlled, it shows that there is an inflection point at which transfusions do occur in anemic patients.

So not being involved with this process before, I'm trying to search is there a process where we could be assured that the panel does have all the data as it makes its decision for what would be a safety issue and a reporting issue at a physician level?

CO-CHAIR CROOKS: Yes?

MS. McGONIGAL. Thank you. Good

morning. Lisa McGonigal from Kidney Care Partners again. National coalition of patient advocates, health care professionals, care providers and suppliers and we work together to improve care for patients with chronic kidney disease.

We appreciate this opportunity to comment again. Yesterday we commented on all of the measure areas except for vascular access, and we're going to use this comment period to address that.

We'd start by saying that we continue our support for the following measures for public reporting only:

NQF measure 0251, which is Functional AVF or Evaluation by Vascular Surgeon for Placement;

0257 is Maximizing Placement of AVF, and;

0259 Decision-Making by Surgeon to Maximize Placement AVF.

KCP continues its support of the

following measures for public reporting, and given the strong evidence that reduction in catheter use has a strong positive impact on fewer infections and hospitalizations and lower mortality, KCP also recommends that the measures be used for payment purposes as well:

NQF 0256, Minimizing Use of Catheters as Chronic Dialysis Access, and;

0262, Catheter Vascular Access and Evaluation by Vascular Surgeon for Permanent Access.

Thank you.

CO-CHAIR CROOKS: Thank you.

Other comments, in person, on the phone? Okay.

So let's go get some food. I presume it's ready. And try to reconvene at 25 minutes to 1:00 for a working lunch.

(Whereupon, the above-entitled matter went off the record at 12:19 p.m. and resumed at 12:38 p.m.)

CO-CHAIR CROOKS: Okay. Let's call

the meeting back to order.

So at this point we'd like to welcome the measure submitters for vascular access to give a brief presentation of your metrics, after which we're going to discuss exactly what order we're going to attack them at. So, shall we start with -- is someone from SVS on the phone?

DR. PACE: Is Lindsey Adams on the phone?

CO-CHAIR CROOKS: Are the phone lines open?

OPERATOR: Phone lines are open.

CO-CHAIR CROOKS: Okay.

DR. PACE: Okay.

CO-CHAIR CROOKS: So if Lindsey is not there yet, let's go to HCQA.

DR. PACE: KCQA.

CO-CHAIR CROOKS: KCQA. Okay.

MS. McGONIGAL: Okay. Thank you.

CO-CHAIR CROOKS: Kidney Care Partners. Please go ahead.

MS. McGONIGAL: Okay. Again, I'm Lisa McGonigal from Kidney Care Quality Alliance, which is an alliance of patient advocates, health care professionals, care providers and purchasers convened by Kidney Care Partners to develop performance measures for ESRD Care.

KCQA care is pleased to submit an information for two vascular access measures for continued NQF endorsement:

Measure 0251, which is Vascular Access Functional AVF or Evaluation by Vascular Surgeon for Placement; and

Measure 0262, Catheter Vascular Access and Evaluation by Vascular Surgeon for Permanent Access.

Both measures were endorsed by NQF in 2008 and they're included among CMS' phase III clinical performance measures. The phase III CPMs are slated for use by CMS in its CROWNWeb dialysis facility data repository when it becomes functional. Both measures

have been demonstrated as reliable and valid through field testing, which was performed both in clinician offices, coincident with the AMA PCPI renal measures and at 53 dialysis facilities across the United States.

The underlying rationale for both measures is to minimize the use of catheter vascular access and maximize permanent access placement in use in all eligible human dialysis patients, as is consistent with the current KDOQI clinical practice guidelines for vascular access and a large and growing body of evidence demonstrating the superiority of permanent access types over catheters.

We note that the KCQA vascular access measures are unique to the NQF renal performance measures portfolio in that they focus not only on outcomes, that is, the percentage of patients with a permanent access, but also on the process of ensuring that those patients without permanent access are seen and evaluated by a vascular surgeon

for placement.

We'd like to thank the Steering Committee and NQF for your consideration of these measures, and we welcome any questions either now or after your deliberations.

CO-CHAIR CROOKS: Thank you.

Representative for CMS?

DR. MESSANA: For the sake of time, we'll not make any major comments other than to remind you all, as you deliberate, that our two measure submissions are linked. That we feel that maximization of AV fistula and minimization of catheters need to be taken as a link set of measures.

Thank you very much.

CO-CHAIR CROOKS: Thank you. Is SVS, Lindsey on the phone now? Okay. We'll defer for a bit. We know they're expected to be on in the near future.

So let's have -- Karen and I sort of had an arbitrary order, but we wanted, before we decided which one to start with we

thought we would ask the Committee, and particularly those who reviewed these metrics if they felt that one or more of them are more important for the Committee to discuss in person today as opposed to possibly being deferred to a phone meeting.

So Andrew already told us that one of his, catheter --

DR. PACE: 256 could wait.

CO-CHAIR CROOKS: Could probably wait because he believes it's pretty straightforward.

Other comments from reviewers?

MS. ANDERSON: It might be good to discuss 0259 Hemodialysis Vascular Access: Decision-Making by Surgeon to Maximize Placement of AVF.

CO-CHAIR CROOKS: Okay. That actually was kind of number 1 on our list for whatever reasons.

So other comments from reviewers? Preferences?

DR. PACE: Okay. Then why don't we go --

CO-CHAIR CROOKS: Well, we can't do that one yet.

DR. PACE: No, we can't do that one yet. But why don't we do one of the --

CO-CHAIR CROOKS: 251?

DR. PACE: Let's do 0251 which is a KCQA measure and Jerry Jackson was our lead discussant.

DR. JACKSON: You want to start with that one? Let's pull it up.

Okay. This measure is: Vascular Access - Functional AV Fistula or Evaluation by Vascular Surgeon for Placement.

The measure steward is KCQA. It's for endorsement. It is a clinician level measure. And --

DR. PACE: Yes, that's right. And just a distinction. The CMS measures would be facility level. This is the clinician level measure.

DR. JACKSON: I believe we were all agreed that the importance to measure and report was high to moderate. Let me look at that specifically.

DR. DALRYMPLE: I apologize. This is Lorien again. I was just verifying the measure we're doing right now.

DR. PACE: 0251.

DR. DALRYMPLE: 0251? Thanks.

DR. PACE: Right. And we'll start with impact, Jerry. So we note the initial reviewers indicated, everyone was in agreement it was high-impact. So maybe we could go ahead and vote on that and then move on.

DR. JACKSON: Yes. All the reviewers agreed it was the same thing.

DR. PACE: Okay. All right. Okay. So we're on 0251: Vascular Access - Functional AVF -- oh, Jerry, I jumped the gun here. Would you give us a description of the measure? I'm sorry. Totally sorry.

DR. JACKSON: Yes, I'm sorry.

Okay. Let me get back. Switching between screens here.

Okay. The numerator is the number of the patients from the denominator who have a functional AV fistula using two needles for cannulation or do not have a fistula with two needles being used, but have been evaluated by a vascular surgeon or other surgeon that's qualified to place vascular access for the placement of an AV fistula at least one time during a 12 month timeframe.

And the denominator statement are all patients aged 18 and over on hemodialysis during the 12 month period who have been on dialysis for greater than three months or 90 days. And there are no denominator exclusions. And the data collection can be from any variety of sources.

CO-CHAIR CROOKS: Okay. So I think we can go to voting on the impact. Any other discussion? All right. Let's vote.

DR. JACKSON: Oh, one other thing

is the steward - I'm sorry.

CO-CHAIR CROOKS: Go ahead.

DR. JACKSON: Listed this is an outcome measure. And I think it's either intermediate outcome or process.

DR. PACE: I think in the past we had these categorized as process measures. But, you know, this is one of those areas where you could kind of look at it in different ways, but I think we've had it categorized as process in the past.

DR. BERNS: If I can just ask a quick question, it doesn't relate to the vote. But on the survey form that was developed that goes along with this that asks whether or not the patient is in hospice. And I'm just curious as to whether that was meant to be an exclusion in the denominator because it's not indicated as such?

DR. PACE: And do you want to answer that right off the top?

DR. NISHIMI: It was a combined

form for the two measures, so that question pertains to the other KCQA measure for an exclusion.

DR. PACE: Okay. So there's no exclusion for this one? Okay.

So let's vote on impact and then we'll get into the more specific --

CO-CHAIR CROOKS: Okay. Voting is open.

MS. RICHIE: Lorien, impact?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: That's 20.

DR. PACE: Okay.

CO-CHAIR CROOKS: All right. Let's do it. All 20 votes were for high impact.

So the next vote would be for performance gap.

DR. JACKSON: Right. Now that, there were two modes of data collection that were carried out at the time of the first submission of the measure. There was a wide cross-section of facilities that were looked

at and then seven MD practices, and they were not overlapping. I'm pretty sure that the MD practices were different than the facilities.

The performance, as judged by the specifications, was 72 percent from the MD offices and 84 percent by facilities. And that was judged to be a gap in performance, although I did not see other data presented that drilled down more to the gap between individual physicians. But there is a gap. And if a 100 percent is the target, than there is a gap in performance.

DR. PACE: Okay. Other reviewers, any comments or other Committee members about performance gap in this area?

MS. ANDERSON: I think my concern was, again, this is at a clinician/physician level. And this performance gap was really done based on facility level review for the most part. And I also feel that the goal of a 100 percent is an unrealistic goal.

DR. PACE: Well, let me just

clarify. The goal is not part of the measure, I think. You know, so, again, it's like we talked about before; performance measures, you know, more is better but there's not like you have to meet a certain threshold.

DR. JACKSON: But if I could interject, I think that comment was based on the percentages put into the application by the developer as representative of a performance gap --

DR. PACE: Oh.

DR. JACKSON: -- my interpretation of 72 percent by the MD practices was 72 percent of what? And I will ask the developer that question. Was the 72 percent of performance by the MD offices based on a projection of a 100 percent, or what's the 72 percent of?

DR. NISHIMI: Two things. The first issue to the point that it was -- this is a facility testing. It was tested in facilities but the level of analysis that is reported

here is to the physician. It was just that the facility's records were used. So I did want to clarify that.

And then did you want to clarify the relative? The question of whether there's a gap compared to what, I mean, ideally yes, 100 percent of people would have some kind of permanent access.

DR. JACKSON: That's what the reported percentages refer to if it were completely fulfilled.

DR. NISHIMI: Yes.

DR. JACKSON: Okay.

DR. PACE: Could you repeat that? We couldn't hear.

DR. JACKSON: The percentages reported in the MD -- in the MD offices of 72 percent and 84 percent in the facilities was based on the ideal of complete adherence to this or 100 percent.

DR. KLIGER: I'm sorry. Just help me. I'm a little confused. Because the

performance gap ought to be measured as those people who didn't fulfill the criteria of this measure. I don't see that data here. Do we have any information on the performance gap?

DR. JACKSON: No.

DR. NISHIMI: The performance ranged from 33 to 100 percent, so -- and we do report that.

DR. JACKSON: Where is it?

DR. NISHIMI: So there is a high degree of variability.

DR. KLIGER: Right. I'm sorry. It's not a matter of which access people have, but whether they fulfill the criteria of these specifications. Do we have that? If we do, I'm sorry, could you just point us to that?

MS. McGONIGAL: No. These are measures of the people who either have the AVF or were seen by a physician, which is fulfilling the criteria of this measure.

The performance in the facilities was 84.4 percent, with a range from 33 to 100

which is substantial variability demonstrating a gap. And mean performance rate of 72 percent within physicians' offices.

DR. JACKSON: Was there a range reported on the MD office data?

MS. McGONIGAL: It was not reported, but I actually do have that data and I could probably dig that up pretty easily.

DR. JACKSON: Because that's one of the things that several of the reviewers commented on, is that the data for facilities does not directly apply to a physician level measure. So we're trying to get to a performance gap by physicians.

DR. NISHIMI: With this particular measure, the data source that's used to report this measure is feasible through facility-based records. Testing in the physician's office required the Iowa Foundation for Medical Care to have both facility and the physician office record. So the best data source for this, to then analyze at the level

of physicians, is the facility's records.

DR. JACKSON: Okay.

DR. PACE: Okay. And the information on opportunity for improvement presented in the submission was based on their collecting data based on the specifications for this measure. And maybe you should just clarify. Because the original measure was not specified necessarily to be collected out of facility records or CROWNWeb data. It was CTP II codes. So maybe it's no longer specified that way, correct, the CTP II codes?

MS. McGONIGAL: We specified it so that, with the intent that it would be collected via CROWNWeb, which would require chart review to enter the data into CROWNWeb. But we also went ahead and specified out codes that we included in the data dictionary. It was not tested using the codes. It was tested using chart review.

DR. PACE: Okay.

CO-CHAIR CROOKS: Can you summarize

what we've learned about the performance gap?

DR. PACE: Well, I think Lauren's got the information up on the screen now about performance gap. And, obviously, the idea is for patients to either have the AVF or to be evaluated for placement. And given that that measure is either/or, the expectation, it should be pretty high.

You know, like all performance gap information, it's relative to the severity of the problem. So the data they presented was that there is variation in performance and overall patients are not always getting either the AVF or being seen by a surgeon for potential placement.

So any other comments about that or disagreement that --

CO-CHAIR CROOKS: And this is in a sample of 1700 patients, so it doesn't reflect national data. And so I'm wondering if this -- has there been improvement? Has this been done serially, and has the gap closed since

the initial endorsement in those facilities or health care entities that use the metric?

DR. NISHIMI: This was originally endorsed under a time-limited status. So the testing was done between September 1st and the end of August 2009. Since then we have not gone out to look at longitudinal data. The published literature would suggest, though, that there still remains an issue with the 72 percent of people or something of that nature starting dialysis with a catheter.

CO-CHAIR CROOKS: Thank you. Okay? So is the Committee ready to vote on performance gap? Okay. Let's do it.

MS. RICHIE: And, Lorien, performance gap? I'm sorry, what was that? Lorien, are you there?

DR. DALRYMPLE: Yes. Can you hear me?

MS. RICHIE: Now I can.

DR. DALRYMPLE: High.

MS. RICHIE: Thank you.

CO-CHAIR CROOKS: Okay. Three votes high, 17 moderate, one insufficient. So we decide this is not an outcome and we should look at the body of evidence, right?

DR. PACE: Right.

CO-CHAIR CROOKS: Okay.

DR. JACKSON: The evidence primarily reviewed four studies. None of them were randomized controlled trials. The evidence focused on the better outcomes with fistulas compared to other types of access, the lower cost, lower complication rate, lower hospitalization and things along those lines.

So the evidence was not precisely aligned with the measure focus, but certainly implied the direction of the measure focus.

DR. PACE: Other reviewers, any comments about the evidence? So the specifics about the evidence was about the lower rate of complications with use of AVF, which is very relevant to the measure.

CO-CHAIR CROOKS: Okay. There were

four studies cited. So we can vote, I think, on the quality.

DR. PACE: And --

CO-CHAIR CROOKS: Okay. All right. So let's vote on the quantity.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: And we'll go with 20. All right. Everyone's getting good at reading and following the chart. Twenty voted moderate.

Okay. Now to the quality. So you mentioned of the four there was no randomized clinical trials, that they support the notion that AVF is good, nothing that was directly studying the metric that AVF or referral to a surgeon is good. Is that a good summary?

DR. JACKSON: Yes.

CO-CHAIR CROOKS: Other comments?

DR. PACE: Other reviewers, Andy or Connie, anything to add about evidence?

DR. FENVES: I mean, the only

comment I would have, of course, part of the measure is to refer to a vascular surgeon. That surgeon may well do vein mapping and decide an AV fistula is not a good choice for that patient and put in a graft. This is still a good clinical outcome, coming from a clinical nephrologist standpoint, but it has nothing to do with fistulas in this case, except that it's a good evaluation of a patient of what is best for the individual. But it's somewhat, you know circumstantial.

CO-CHAIR CROOKS: Does the specification say that the surgeon has to have a plan for an AVF or just that the patient be evaluated or just referred? Evaluated?

DR. JACKSON: I was going to get to that under reliability and specifications. But it's documented in one of four ways. The nephrologist can dictate a note into the patient's chart, the surgeon can dictate a note, the staff member at dialysis can dictate a note. And then, if the surgeon chooses for

whatever reason not to place a fistula, that reason needs to be documented in the patient's chart. So there's those very specific specifications that allow that to occur.

DR. PACE: And just a little bit of history. The last project where this was reviewed, there was discussion about referral. And the Committee really strongly encouraged, and the measure was modified at that time to actually include evaluation, not just that there was some referral --

DR. JACKSON: Intent to refer?

DR. PACE: Right.

DR. JACKSON: So that word has been changed.

DR. PACE: Right. Right.

DR. JACKSON: Yes. It's actual evaluation.

DR. PACE: Right. Shall we move on?

CO-CHAIR CROOKS: Okay. So are we ready to vote on the quality of the evidence?

Let's do it.

MS. RICHIE: And, Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Nineteen moderate, two low. Okay.

And consistency, Jerry, any thoughts, advice?

DR. JACKSON: I think the consistency that fistulas are better than anything else is high. I mean, the relationship of the evaluation by the surgeon component of this is not well studied. So I think, you know, how does that change? I'd like for other people to comment about how does that change the assessment of consistency.

CO-CHAIR CROOKS: Well, I --

DR. NARVA: That was my concern, too. I mean, I think this is a very strong case for obviously having fistulas but this is not a strong case that this process -- that the behavior that's mandated in this measure

is going to lead to that.

DR. JACKSON: For instance, just drilling down a little bit, the process varies by location. But for the most part, our surgeons want a mapping done prior to them seeing the surgeon. Sometimes that mapping indicates something different. It might affect where they go. So it's going to be done a variety of ways in different places, but obviously, evaluation by a surgeon, whatever that means, has to occur before they place a fistula. So I'm not sure that that is that germane to the consistency question.

CO-CHAIR CROOKS: Alan?

DR. KLIGER: I think this is asking us a question about the consistency of the data, not of our process. So you've already said you think the consistency of the data are high.

We do need to talk some more about the process, and perhaps unintended consequences of this.

DR. JACKSON: Fair enough.

CO-CHAIR CROOKS: Okay. So can we vote on consistency? Any objections? All right. Here we go.

MS. RICHIE: And Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: Seven votes for high, 14 moderate. So it does pass the evidence decision logic grid with a yes. And so we don't need this.

And we did meet all three subcriteria, right?

DR. PACE: Right.

CO-CHAIR CROOKS: Okay. We don't need this one.

So measure properties, reliability and specifications?

DR. JACKSON: On reliability, there was, I think, a very high level decision in the application. They had gone back and done data integrity audits in 11 out of 53 sites that I think were at facilities. And then in

both the MD offices and facilities there was inter-rater reliability that was assessed that had high kappa scores. So at that level of reliability, I personally thought that was impressive.

In fact -- can I talk about specifications right here?

DR. PACE: Yes.

DR. JACKSON: One major issue I had with specifications is that because of the problem with increasing catheters and other issues, there's been a slight upward blip in the prevalence of grafts. For patients who have a graft that is functioning well or even who has an occasional intervention according to KDOQI guidelines, that person would not need evaluation for a fistula as yet. It would be when the graft starts failing or has, I believe, three interventions within a six month block of time that they would need to be referred for evaluation for a secondary fistula.

So, I think there's an issue in the specifications with -- leading to the unintended consequence of overuse for the approximately 15 percent of people who have grafts that are functioning well that would be required by this to see a surgeon annually for fistula evaluation. So, any comments from Connie or others?

MS. ANDERSON: I agree with that. I think there's another unintended consequence and it's for those patients that have catheters that have been evaluated by a surgeon and have been deemed to have access never, meaning at no point will they be able to have an AVF or an AVG. And so, again, the burden of those people having to be evaluated by a surgeon when it's been deemed that they will not be able to have a vascular access of AVG or AVF.

DR. JACKSON: I suppose there could be a specification requiring a second opinion in that case.

MS. ANDERSON: Or have them as part of an exclusion criteria.

DR. JACKSON: Right. Hospice would be another like that. That would be another exclusion.

MS. ANDERSON: Yes.

DR. PACE: So it sounds like we're getting into some validity issues with how it's specified.

DR. JACKSON: Right.

DR. PACE: And I know that this seems like splitting hairs, but just to help us kind of keep things in category and give the right feedback, the specifications, as they are, are pretty precise. And you indicated the reliability. And then I think this is good discussion that we definitely need to bring into the validity question.

DR. VELEZ: But don't you think, Jerry, I mean, what I heard you mentioning is also what Andrew mentioned earlier, is: should we have exclusion if they have a working

graft? That's what I heard you say.

DR. JACKSON: Should we go ahead and vote? We'll come back to that when we talk about validity.

DR. PACE: Right.

DR. BERNS: I do have one question that may relate to this, but tell me if not. And that is the definition of a surgeon qualified in the area of vascular access and whether that is something that -- the reliability of that assessment was determined? You know, in other words, that's a judgment call that may or may not be correct.

DR. PACE: Right. How is that defined is your question, right?

DR. BERNS: Yes.

DR. PACE: No, it relates to precision of the specification. So we can ask the developer if they have a definition for that or how they --

MS. McGONIGAL: Yes. I know that the measure was originally specified that way

to address the issue of remote areas where there would not necessarily be a vascular surgeon present. And in those situations, it would be unfair to not give credit if a patient was referred to a surgeon who does do the vascular access for that area. That was the rationale behind writing it that way.

DR. BERNS: I would argue the flip side, that there are vascular surgeons who are not qualified to do vascular access.

DR. DALRYMPLE: I had a question about the data field. Are all of these data elements on page 9 going to be included, or is this a combination of using CROWNWeb and chart reviews? So for example, note or letter prepared by the nephrologist or the personnel --

MS. McGONIGAL: All of the access types are a part of the CROWNWeb data fields currently. CROWNWeb does not currently have a data field for seen or evaluated by a vascular surgeon. However, we have been in discussions

with them and they have indicated their interest in including this measure with the next iteration. How they will go about including that data field, we're unable to speak for them at this time. But they do intend to do so.

DR. BERNS: Let me just return to this point. This is a subjective component of this which is unusual for these performance measures. So I'm not convinced that the wording about appropriate or qualified is really appropriate for this kind of performance measure, because it confers, then, an opinion as part of the performance measure that that surgeon is in fact qualified.

DR. PACE: So would a solution be to just say to a surgeon -- I mean -- and not realizing that --

DR. KLIGER: How about interventional nephrologists that do this? They're not surgeons.

DR. PACE: Oh.

DR. JACKSON: Well, this gets to my question about --

DR. PACE: So, could you leave out "to whom" and say "evaluated for placement"? I'm just --

DR. JACKSON: Well, I think the goal is to get a fistula in as high a percentage of people as possible. And especially for catheter patients I think it is very necessary for them to be evaluated by a surgeon who is capable of putting in a fistula. And, you know, there's been a lot of type of small volume writings in the literature about the scope or the range of surgical abilities. And there'll be many surgeons, maybe a majority, who would say a patient could not have a fistula and in fact will have several grafts that fail, and then eventually another surgeon who is higher skill level operator for fistulas will put in a very well-functioning fistula. So it's extremely subjective when the patient sees any surgeon,

whether qualified or not, whether or not they're eligible for a fistula. But I don't know anyway to get around that.

DR. FISCHER: Well, why not do, Karen, as you suggested. Because the processes of care may be variable depending on one's setting, whether it's a transplant surgeon, a vascular surgeon, general surgeon or interventional nephrologist. If you just say "evaluated for" -- I just wonder if that's a reasonable way. Because I don't know -- all of us may operate in different care settings and how that goes about may be highly variable and be very difficult describe in all those details into this.

DR. BERNS: It might be reasonable to phrase it "patients seen or evaluated by a vascular surgeon or other physician for an AVF." Then that would get to the nephrologist issue, it would get to any type of surgeon.

DR. PACE: Okay. So where we're at with -- generally we do this based on the

measure as specified. I guess we could ask the developer if they would be amenable to that language or if we should -- or maybe we'll just vote on it as it is and then we can see if there's a recommendation that comes to you. Well, let's do that.

DR. NISHIMI: Yes. I was going to say it struck me that you should first vote on it and then recommend what you would like to see.

DR. PACE: Right, right. And that's what we've been doing. So I won't interrupt that process.

CO-CHAIR CROOKS: Okay.

DR. PACE: So we're talking about voting on reliability and this includes precise specifications and reliability testing.

CO-CHAIR CROOKS: Okay. Any other discussion? Shall we vote? Okay, let's vote.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: The result: 17 voted moderate and four low. So it passes reliability.

DR. PACE: Okay. All right.

CO-CHAIR CROOKS: So, keeping the discussion in mind for later, let's go on to validity.

DR. JACKSON: The developer spoke to a process of emphasizing how the sites chosen were highly representative of the broader populations. So the sites were well selected and statistically tested for representation. The question arose in some of the comments as to whether that was a valid testing of validity, essentially we're talking about validity.

And then that aside, face validity was referenced but not in a -- what we talked about here is a systematic way. The committee doing that was not listed.

And then also the previous endorsement process, the CDP was referenced as

face validity, which I'm not sure we'd accept.

Comments?

DR. FISCHER: I had one question about -- well, can we talk about specifications as well as it pertains? I think I'm just following up on comments.

Just rereading this -- so I just want to make sure if I have a patient who their prevalent access, they're working access is a graft or a catheter and they've been on dialysis for, let's say, ten years. And they were evaluated two years ago or this access was placed eight or nine years ago, it's been working fine. I mean, this says a 12 month reporting period. So if they had been evaluated previously outside of the 12 month reporting period and were deemed not suitable for a fistula, and therefore a catheter or graft, then they would not meet this measure or is that not the way? Because it seems like the 12 month reporting period, then every year I have to have them go back and see a surgeon

when we've already kind of been down this road.

DR. LATTS: Then you need to put in the exclusion that Jerry mentioned earlier for the well-functioning --

DR. JACKSON: Okay. A well-functioning graft. I think if they have a catheter, that's a little different situation, especially as surgeons have learned better ways of doing translocation and transpositional fistulas, et cetera. So the skill level has improved. Certainly catheters are a high risk, but if you just look at KDOQI guidelines and common practice if someone has a well functioning graft without problems, and -- do they need to see a surgeon year after year prior to the time that the graft fails is the point. It's probably about--

DR. NISHIMI: If I can just address the issue of graft, I might be able to short circuit this conversation a little bit.

As the developers, we tested the

measure as it was originally endorsed. But we also gathered information on permanent access broadly, i.e., with grafts. So we have the data and we would be very amenable to a recommendation from the Steering Committee, not to exclude people from this measure, but to redirect the focus of the measure to be functional permanent access, if you will. Or whether permanent access or if you got a catheter, you need to be evaluated.

DR. JACKSON: Our discussion.

DR. NISHIMI: Right. So it was tested as it was endorsed, but we recognize that the shift towards grafts -- so we collected that information. The reliability information that you see here is really no different. And we would be amenable to having you recommend it. So that means that the measure could more accurately reflect appropriate practice.

DR. FISCHER: I don't want to just -- I mean, I have -- there are circumstances

where a catheter may be the patient's only option. And I don't want to get on that too much, but I just -- I can cite two examples right off the bat. Patients with congenital heart disease frequently if they're transitioning from pediatric to adult populations. Some of them will develop high output heart failure with permanent vascular access. The second case is patients who have behavioral cognitive problems who will not tolerate having two needles in their arm.

So, I just think that there are clinical circumstances that do occur. I mean, I don't know how that would be accommodated in this measure.

DR. KLIGER: Mike, I thought you might say something like that. Because I really want to underline this. I think one of the unintended consequences of the fistula-first project was to really ignore patient choices and patient stratification by need.

We know that in the best of all

worlds the fistula probably is the best access. But for individual patients it might either be impossible, impractically or clearly not the patient's choice, for whatever reason.

I can tell you in our FHN study when we looked at our home patients doing six times a week at home dialysis, a substantial portion of those patients used catheters. And we're going to be discussing the vascular access issues at the ASM coming up. But I can tell you that the catheters are not so bad for those people and the complications are not the ones that have been described before.

So, I'm just very concerned that what started off here as an overall recommendation about the best type of vascular access that we've learned since then about potential variation and potential patient-centered care that make me concerned about the measure.

DR. KLEINPETER: One other thing, looking at some of the older patients, I think

it's really cruel to send those people to surgery over and over again, particularly those that are starting dialysis above the age of 80. And we go straight to graft in my program and they have just fine outcomes. They're not in the hospital constantly. And there needs to be some type of consideration for some of those other older patients.

DR. NISHIMI: Well, and that's why we're amenable to the Committee recommending that graft be encompassed by this measure --

DR. KLIGER: So, I'm saying more then just graft, I guess.

DR. JACKSON: It sounds like when we get to the useability we need to recommend some exclusions as well. But --

DR. PACE: Well, I think what we would need to do is vote on validity and if these issues about the specifications and whether that makes it a valid indicator of quality, and then if that's the reason -- if it doesn't pass this and that's the reason,

then we can make the recommendation and they can come back to you all with that change specification. Would that make sense?

DR. JACKSON: Could I reframe what I said earlier as a question to Karen? And that is, the methods of validation or validity in the application, do they meet with NQF guidelines validity?

DR. PACE: The discussion about the characteristics of the study sample are not exactly what we're looking for validity of the measure or validity of the data. That certainly provides good evidence about the method that they used for testing the reliability.

In face validity we ask for a systematic assessment. And, you know, I think that's something that you all can judge. You know, the fact that it went NQF endorsement. I mean, what the task force I guess had in mind was more about new measures. So I don't think they had necessarily considered that as

one of the things that people would present.

I think you all can apply your own judgment to face validity as well, or again that's something that we could ask them provide us some information on.

DR. JACKSON: And if I misled, I'm sorry. There was a panel separate from NQF's CDP, the members were just not specified in a way that we've had on other applications. So three was a panel and it was stated that the panel accepted this on face validity. And I believe that's right. Yes. So there was some level of validity, it's just that with the new guidelines --

DR. PACE: Right. The new task force guidelines is that they were recommending that we get more of a systematic assessment of that face validity. But, again, on face validity I think you can either ask for them to do that or kind of go on your judgment of face validity.

MS. McGONIGAL: Karen, I just

wanted to add that we did include all of the names involved in the expert panels that were involved in overseeing the development and approval of these measures down under "Additional Information."

DR. PACE: Okay. So, let's just go to that.

MS. McGONIGAL: Page 22.

DR. PACE: Right. Okay. Any other discussion, questions, clarifications?

DR. DALRYMPLE: So, Karen, are we supposed to vote on that measure as -- and then if it does not pass, would there either be an opportunity for us to make some votes that have been discussed?

DR. PACE: Yes. Yes. In a minute we'll vote on the measure as it currently stands as it's specified.

MR. WELLS: I think when I evaluated this measure, and I might have been mistaken, probably was. But when I read the 12 months, those greater than 90 days, I guess in

my mind I was thinking of those that just initiated dialysis and had to be seen within that time period. And I guess when I look at the validity of it, I guess I just take, you know what I read in there. And I mean it just seemed pretty straightforward to me. I didn't drill down to, you know to evaluate all the -- I guess the exceptions or what have you.

And I think the number of exceptions to this, the elderly and what have you that wouldn't be suitable for a fistula, I think that's going to be a very small portion. And I think to me the initiation of a fistula is very important. And, you know I was very fortunate when I got mine. I mean, my doctor, I mean I don't think he wanted a catheter in me anymore than ten days. But my fistula didn't become functional until about four or five months after it was placed. So I had catheter for a pretty long time. And one of the happiest days of my life was getting that thing out.

DR. PACE: And that is a point about exclusions. I mean if it's a very small number, then again it's probably more documentation and data collection burden that contributs to the measure.

CO-CHAIR CROOKS: Jerry?

DR. JACKSON: When we vote on validity, can we -- I know what you said about voting on what's in the application. But since the developer's already accepted working with us to take functioning grafts out and do a specification modification, could we include into that in the consideration for voting?

CO-CHAIR CROOKS: What we're voting on is validity as presented here.

DR. JACKSON: Okay.

CO-CHAIR CROOKS: For this metric. And --

DR. PACE: And then --

CO-CHAIR CROOKS: -- even though it's not perfect or there's a lot of considerations, you know we're going to vote

on it as it is here. And then we'll have the opportunity if we think there's ways it could be improved or things they should consider, we can make those recommendations.

DR. PACE: Right.

CO-CHAIR CROOKS: Right? So, are we ready to vote? Okay.

MS. RICHIE: Lorien?

DR. DALRYMPLE: As currently stands, low.

CO-CHAIR CROOKS: Okay. We have 21 already. Okay. So we have eight people voting moderate and 14 voting low.

DR. PACE: Okay. So let's then see if someone wants to propose a modification to the specifications. And what we can do then is vote on that and ask the developer to come back with those changed specifications. So--

CO-CHAIR CROOKS: So start with the largest flaw is the grafts should be included in the numerator and denominator, functioning grafts.

DR. JACKSON: As long as they're functioning well and do not fall under KDOQI guidelines for --

CO-CHAIR CROOKS: I'm sorry, Jerry. I'm not hearing you very well.

DR. JACKSON: As long as the grafts are functioning well and do not fall under the KDOQI guideline for referral for a new access based on frequency of intervention.

CO-CHAIR CROOKS: We'll consider that. That may be hard to get into a data form, you know. The last --

DR. JACKSON: Just like --

CO-CHAIR CROOKS: -- used access was a fistula, that might be as good as we can get, something like that. But anyway, this is advice to make the metric more acceptable and valid for us.

Other suggestions to put on the record?

DR. BERNS: We talked about hospice. We talked about elderly patients.

CO-CHAIR CROOKS: Hospice patients.

DR. BERNS: Patient choice where -- you know, at some level this is out of our hands. You do all you can do and the patient says I've been dialysising with a catheter, and my neighbor died with a catheter, and my neighbor bled out from their fistula or whatever, and I'm not going to go see the surgeon. Or they go to the surgeon and they never get the follow-up appointment to get the surgery performed. So the physician has done everything right and yet there is still a significant number of patients who will never end up getting an AV fistula. And I'm not sure how you can --

CO-CHAIR CROOKS: Well, I'd just like to comment on -- and having done a lot of QI on vascular access, patients who don't want -- just want their catheter, you know, I think we want to not institutionalize a system where you just let it go at that, you know. That often reeducation, bringing the issue again,

sending them to the right surgeon.

I mean, some surgeons will look at a patient and say, "No way, I'm not even going to try a fistula." And another one will say, "Sure. I just need a venogram, here's the place. Boom it's in."

So I don't think we should -- except maybe in the case of a hospice patient, a patient with a very short life expectancy who does not want the inconvenience of a surgery, maybe you could come up with very few other -- and maybe a patient who just cannot risk any increased cardiac output for any reason. Other than that, I don't think we should exclude.

DR. BERNS: Okay.

MS. ANDERSON: I do think the other exclusion is those that are already being evaluated by a cardiovascular surgeon or a vascular access surgeon and the surgeon deems them unsuitable for either an AVF or an AVG.

CO-CHAIR CROOKS: Well, again, I'm

a little hesitant there for the same reason. There's different surgeons. But consider that.

Also, we were worried somewhat about the one year time horizon. In other words, if a patient was evaluated a year ago and there's a plan for a fistula, you know, when the graft fails or they're not ready to have the fistula put in yet but they've seen a surgeon, do they need to go back in 12 months? Alan?

DR. KLIGER: Well, Peter, I've heard some difference of opinion around the table about this. And it seems to me we're not going to resolve this but simply that we need to ask the developer to have heard all of these discussions and arguments and then to consider what they want to do.

CO-CHAIR CROOKS: Okay. Yes. I think we've stated into the audiotape all--

DR. NISHIMI: Yes. I mean, we're cognizant of the discussion. I think we know

what we can do within the data that we have. And we'll come back to you with a revised measure.

DR. PACE: And just one other comment about the -- you know, we do have a facility level measure that's just about AV fistula and we don't have all these exclusions. And we do need to think about that, again, what's the frequency of these exclusions, what's the differences in distribution? So it's probably a measure that you're not going to get at 100 percent or zero percent, but it's that you have fair comparisons. So we can ask them to address those and come back to you with some analyses and changes.

Jeff?

DR. BERNS: It may get to the point that you mentioned about frequency of exceptions. But the definition of functional fistula really only requires one occurrence with two needles, as I read it.

As you're thinking about revising it, you may want to think about revising that part of the definition as well.

CO-CHAIR CROOKS: Okay. So I think we can leave this metric now and move on to another.

And let's see if SVS is on the line. Lindsey or another person?

DR. XENOS: Yes. Hi.

DR. KRESOWIK: Tim Kresowik is on too.

DR. XENOS: Yes. And Eleftherios Xenos.

CO-CHAIR CROOKS: If you're not picking up your handset, please do that. You're coming across kind of distorted.

DR. PACE: And could we have one of you give a brief introduction to your measure? This is would 0259.

DR. KRESOWIK: Yes, I can do it if you want. This is Tim Kresowik.

The measure is basically -- I've

listened to the last discussion, but it's basically the surgeon's counterpart of patients being referred for vascular access with the concept that -- to encourage fistula over graft. And again, I'm well aware of all the controversy there. But with the exception that it's based on vein mapping and the specifications really do allow more than that it terms of physician exclusion based on their judgment that the patient is not a candidate for an AV fistula.

So, I mean, it's a pretty simple concept and a relatively simple measure.

CO-CHAIR CROOKS: Okay. Thank you.

The reviewer is Connie.

MS. ANDERSON: This measure is the percentage of patients with advanced chronic disease, CKD 4 or 5 or ESRD undergoing open surgical implantation of a permanent hemodialysis access who receive an AVF.

The numerator is the patients undergoing hemodialysis vascular access

procedure who have an AVF or who receive an AVF. And then the denominator is all patients with CKD 4, 5 or ESRD who have surgical placement of permanent hemodialysis access.

So this is a process measure and it's at the clinician level.

In terms of impact and importance to measure, I think it was pretty unanimous that this is a high impact and that AVFs have the highest long term patency rates and lower rates of infection. And so there's a high impact in order for this measure.

CO-CHAIR CROOKS: Okay. Shall we carry through the discussion about the high --

MS. ANDERSON: And I think --

CO-CHAIR CROOKS: Alan?

DR. KLIGER: I'm sorry, just before we get there the box of whether or not this has been tested or not is not marked. And so if it's untested, as I understand it we're not going to be discussing it. Do we know whether it was tested or not?

DR. PACE: Is this one we --

MS. RICHIE: I think this is the one that we don't have that information.

DR. KRESOWIK: It was submitted previously.

This is Tim Kresowik again.

I was not involved in that testing process, but it has been previously submitted.

DR. PACE: Right. So I think that's a good point and we probably can't continue discussing it at this point.

Did you look at the -- let me just look. No. Go to 2.A.2.3. There's some data. That was probably checked incorrectly. There's some reliability testing data.

MS. RICHIE: 2.3. It's on page 7.

DR. PACE: And validity. And it's basically the CPT and the ICD-9 codes. And there were, it looks like, two practice groups. Yes, so we can go on and then we'll evaluate that data. Okay.

So, impact, is there any other

discussion about impact? Should we vote on that and then go on with the other thing? Is that okay?

CO-CHAIR CROOKS: Okay. Let's vote on high impact. On the impact: High, moderate, low and insufficient. Starting now?

MS. RICHIE: Lorien, impact?

DR. DALRYMPLE: High.

CO-CHAIR CROOKS: That's 21. So 17 high, three moderate and one low. Okay.

Onto the performance gap.

MS. ANDERSON: Currently based on the data presented, which was April of 2010, there's a 55 percent rate of AVFs with a goal of a 100 percent. So demonstrated performance gap.

DR. PACE: And we should mention this is a previously endorsed measure.

MS. ANDERSON: Yes.

DR. PACE: So it's up for endorsement maintenance. And did they provide information on the actual measure?

CO-CHAIR CROOKS: Or that Fistula First is the same measure?

MS. ANDERSON: The Fistula First is where they gathered the data from.

DR. KLIGER: Right. But the measure -- I'm sorry, but I guess -- I understand the data on fistulas, but the question is of all people who have open procedures have they looked at how many have these measured? Because that's really what we're asking here.

DR. PACE: Right. So the developer, I know you've tested the measure. Is there any other -- there's no implementation of this measure yet, is that correct? So the only data specifically on this measure is what's in testing, is that --

DR. KRESOWIK: Well, it has been implemented through PQRI being transitioned to PQRS. But we don't, as you all know, CMS does not release the national data for us to be able to analyze that. But it has indeed been

implemented.

DR. PACE: Have you tried requesting that from CMS?

DR. KRESOWIK: I don't know that we've done it in the last few months. I know it's been done previously on other measures. But unless they've changed their policy, it has not been necessarily possible to get the -- and again, if you think about the way the measure is structured with the exclusions, I'm not sure that's going to answer the exact gap question. Because -- in terms of the possibility for improvement, which I think is still based on that current literature.

DR. PACE: Okay. So this is an endorsed measure with no specific data other then the testing data. But that's the case with some of the other endorsed measures we looked at. So, you know, the key issue is is there still opportunity for improvement in this area?

CO-CHAIR CROOKS: Well, we do

Fistula First data is a similar metric, although this metric takes out catheters. And it's not the same as prevalence under Fistula First, which is prevalence of all three types of vascular access, where this is saying if a vascular access is created, what percentage are fistulas and what percentage are grafts. But we do know that there is still a gap. That there's -- many more fistulas could be created. I think we know that from AV First.

Jerry?

DR. JACKSON: If I'm reading the specification right, any patient the surgeon feels that's not a candidate for fistula is excluded. So that includes graft patients, I think.

DR. KRESOWIK: Correct. And the key part of the specifications is that you have to have documented a specific reason why a fistula is not being placed. In other words, if you're putting a graft in and the most common would be inadequate vein based on

vein mapping. But it does not specifically say that that's the only reason.

DR. KLIGER: So let me just -- maybe the developer can help me. This feels a little confusing to me.

If the surgeon says, no, fistulas are not possible here and those patients are not excluded. So the only ones who are included are those for whom the surgeon in advance think the fistula is possible. This then measures the correctness of their pre-op assessment?

DR. KRESOWIK: No, it really doesn't. I mean, this is very similar to a lot of other process measures that are currently in use, which is, you know, basically just looking at the denominator of patients who are undergoing the procedures. So the exclusion has to be specifically designated, okay? So that means a choice. Someone's got to go and say, you know, "I understand that a fistula should be placed. This is the reason I'm not."

So in the denominator if no exclusions are, if you will, included or you don't exclude anybody, they will still be in the denominator regardless of whether you put in a graft or fistula. Am I making that clear?

DR. FISCHER: But it seems like then that this would be 100 percent, is that not --

DR. KRESOWIK: Well, it should be. I mean, yes, it should be if you're --

DR. FISCHER: I mean not to be flippant, but it seems like if -- because the options -- if you're undergoing an open procedures, I only know of two options, a graft or fistula. And if we exclude people who aren't fistula candidates based on -- I mean, this is fine, but I'm assuming that there's going to be high performance on the measure in general, but maybe I have a misunderstanding. But I think that's kind of what Alan might have been asking.

DR. KRESOWIK: Right. No, I understand. And I think -- I mean, this is probably not the time to go on a whole discussion about the optimal way to do measures, but I would say that almost every process measure out there that allows patient or physician level exclusion could receive the same criticism, you know, in terms of the performance should be at a 100 percent if the physician is thinking about it, documenting their rationale.

And I guess, you know, the counter is to try to -- just listening to the discussion that you all just had about all these possible other exceptions and the kind of perverse incentives if you don't allow these kind of exclusions of where you end up with -- you know, you have a potential for doing harm with the measurement. But I'd be the first one to say that, you know, and it's true for most of these process measures, in terms of, you know, certainty that the right

thing has been done. There's just no way to do that.

DR. JACKSON: Let me try to rephrase Alan's question to the developer.

If I understand this correctly, it's testing the success rate of the surgeon in putting the fistula in if he or she up front feels that a fistula should be done. But the problem is that the subjectivity at the start such that if it looks like it's going to be dicey to get a fistula in, they could just say it's not possible and they're excluded.

So my question would be: What is there to keep this from just becoming a slam dunk kind of measure for the surgeon? You know, they're still going to have some OR failures where it just won't go, and it'll measure that. But it looks like it's going to be 90/95 percent for any accomplished surgeon. Am I missing something?

DR. XENOS: Yes. Actually, that is

not true. The rate of non-maturation of surgeons' fistulas have been shown closer to the 30 percent range.

DR. KRESOWIK: But in terms of the question, you are correct. And I think what we're trying to say and similar to, again, going back to the discussion you just had, the alternative is a very perverse incentive. Okay?

As a surgeon, I mean, I can create a fistula in anybody that has almost no chance of success and meet a measure, charge Medicare and then come back and finally have to put a graft in or leave a patient with a catheter. For example -- I'm taking it to the extreme.

So, the alternative is either to not accept those types of exclusions where someone's made a reasoned judgment versus to have a crude measure that just says what's the percentage of fistulas. And then you get into all the, as I said, the perverse incentives, the variation in practice in terms of what

kind of patients are being referred, et cetera.

We're certainly open to suggestions about how to do this better, but I'm not sure how to.

CO-CHAIR CROOKS: Well, I'd like to take a shot at putting it in the paradigm I think the surgeons look at it from.

This metric offers a surgeon a chance at a 100 percent if they either decide and successfully place a fistula or they carefully evaluate whether a fistula can be done and they decide no. Where they fall down is if they don't consider the options, document their decision process and then they go in and put in a graft. That's where they fail. Do you see what I'm saying?

So from the surgeon's point of view they have the chance to score a 100 percent and it sort of forces them to think about it, a fistula, and to document it if they don't think they want to do it.

Jerry?

DR. FENVES: I think it's also worth pointing out that there's no requirement that the fistula mature or ever be used. It's just create a fistula, which is what we've run into as being a lot of the unintended consequences of the last several years.

CO-CHAIR CROOKS: But this may allow them a way out so they're not forced to put in fistulas that they don't think are going to succeed.

DR. BERNS: Put in a fistula whether it succeeds or not.

CO-CHAIR CROOKS: If they don't think it's going to succeed, they can write a note saying this is not a fistula candidate, and not they still score on the metric.

DR. PACE: The metric also doesn't require that it be a functioning fistula.

CO-CHAIR CROOKS: Right, it doesn't. I mean, that's true.

DR. KRESOWIK: Yes. I think what

we're getting into would require -- in fact, we're working on this in other areas, but really getting to true outcome measures. But that's sort of a different step. This is an endorsed process measure and we're rapidly working on other measures that will be better and true outcome measures. And that could be something to definitely work on down the line. But we're not there yet, and this is sort of a separate issue.

DR. FENVES: Can I just have a point of clarification? I think somebody mentioned the word 30 percent non-maturation rate. Did I hear that correctly? Because I think that's truly incorrect because the largest study that was since this measure was approved published in JAMA in 2008, that that fistula study indirectly showed there was 60 percent failure rate in both the placebo group and -- it was a very large study, over -- I forget how many patients.

Now I don't know if you believe

that, but that was a prospective randomized study. And the failure rate was 60 percent.

I should say, we should also maybe piggybacking on what somebody else said, of useability. I should really make that point. Because, yes, there were fistulas in, it's just they didn't work. I mean, there were doppler sounds, but they couldn't be used. And so that's another issue. They could never have two needles placed.

DR. XENOS: Yes, and I agree with that. I mentioned that number, and I should have said at least 30 percent. You're right about that. It might be more. But the lowest number I've seen is 30 percent.

DR. KRESOWIK: But all of those arguments, though, would argue for the measure the way it's specified and include the exclusion. Because otherwise, again, you have that perverse incentive of just putting a fistula in no matter what to get your quality check, if you will, regardless of whether

that's ever going to be used by the patient or useful at all.

So, I mean, I think that is exactly the reason why the specification is as it is.

CO-CHAIR CROOKS: Alan?

DR. KLIGER: I guess my problem is without actual data -- or should I stop? Sorry.

CO-CHAIR CROOKS: I was -- meant to call on Ruben, because he was first. And my finger just automatically goes to Alan every time. I'm sorry.

DR. KLIGER: All right. I hope you can understand my accent. It's a Puerto Rican -- no.

I guess my problem is without actual data to review this metric to see what that really has looked like, it's very hard for me to know if there's really a performance gap that matters or its useability. I surely feel -- what I hear the developer discussing makes real sense in terms of finding the right

way to incent vascular surgeons to put fistulas as often as they can. But without being measured, it's very hard for me to know whether it accomplishes that or not.

DR. KRESOWIK: Yes. Part of the problem, and if you just think this through a little a little bit, this measure is implemented through PQRI. And if you looked at PQRI across the board for all the measures that are being used in there, the performance rate is very high for all kinds of measures. But that doesn't really tell you whether or not a performance gap exists. And if you only use that data, you're going to vastly overestimate performance. Because under a system where you have voluntary choice, voluntary reporting, people of course are -- the early adopters are the ones that are actually doing this, are going to pick things that they're going to have a high success rate and they're going to make sure they have a high success rate.

So I'm not sure that that data will really tell us whether or not there is a gap. And so you have to turn to more or other data sources to really decide whether or not there still exists a performance gap across the country that this measure could address if it was more widely adopted and used. Does that make sense?

CO-CHAIR CROOKS: Thanks. That makes sense.

Ruben, did Alan speak for you or do you have something?

DR. VELEZ: Thank you, Ruben.

I think we now understand what this measure asset is -- measures. But at the end of the day I'm not sure if this information helps us, and it says more to the developer. I'm not sure it's going to help us in achieving what we want to achieve in the outcome. As has been well stated, the percentage may get quite high because of the numerator or the denominator.

DR. KRESOWIK: Agreed. And, you know, again, I would only say that we are in the process of across the board in vascular surgery of trying to develop true outcome measures that will ultimately get us where we want to get for a lot of areas across the board in medicine. But I think if we really look at what's going on, what's endorsed out there right now, the vast majority of them are process measures that all have these kinds of limitations in terms of getting us to where we want to go.

DR. PACE: Just one thing we've been conferring a little bit about, and I think it's a good point of some of the issues about how the measure is constructed and then not having any data to know that plays out and whether the measure is really going to ultimately tell us something. And we understand that everyone's had trouble getting PQRI data from CMS, but something to think about is whether we want to suspend things

here and make it a request to get some actual data on this measure and see with this is kind of holding things up with NQF endorsement, whether that can help get some data from CMA. I don't know. And I guess we could also see whether that's going to -- you know, if you want to go ahead and vote on this performance gap with the information you have, and then we'll see where we're at after that.

CO-CHAIR CROOKS: I would point out to the Committee, if we vote and the result is insufficient data to judge the performance gap, that stops it at this point. And then they can take that under advisement and go from there. Personally, that's what I'm feeling right now. There's insufficient evidence to judge whether there's a performance gap. Vascular surgeons, between the two options, maybe hitting 90/95 percent. I have no way of knowing.

DR. KRESOWIK: But again, I would assume that if we were able to get the PQRI

data, it's going to have very high performance. But that shouldn't be used -- I don't think the PQRI data is the valid way to assess a performance gap. The performance gap has to come from other sources.

DR. KLIGER: Right. So we have insufficient data. I think that's really what you're saying. We have insufficient data to judge a performance gap.

DR. KRESOWIK: Well, why isn't the Fistula First data which shows still a relatively high percentage of grafts versus fistula --

DR. PACE: Right. This is Karen. Let me just explain. The difference is that in general, yes, I think the group agrees there is room for improvement about placing fistulas. What we're addressing here is endorsement of a specific measure and how its specified. And if this measure doesn't really help us identify differences in quality across providers, it's not that useful from a quality

performance metric.

I think -- does anyone else want to add to that?

CO-CHAIR CROOKS: And also the Fistula First information, which is improving rapidly even without NQF direct involvement, but -- is not the same metric. It's a lot different than what this is. And it's true that your performance measurement will be in a limited group of surgeons, I presume, but -- in itself if you explain why if the gap is low, it still may not be accurate. But nevertheless, we need to see some performance data on this metric.

So I think we've finally reached a point where we can take a vote, unless anybody objects. Okay. So let's vote on presence of a performance gap; high, moderate, low, insufficient.

MS. RICHIE: Lorien, performance gap?

DR. DALRYMPLE: Insufficient.

CO-CHAIR CROOKS: Okay. We have 18 voting insufficient and two low.

So I think also in the interest of time we should stop consideration of this metric at this point.

Is it true, Karen, that if they were able to loosen some performance data out of CMS and get it to us within weeks, we could still look at it or -- ?

DR. PACE: Yes, I think so. And so given that potential scenario, do you want to evaluate the evidence or just wait and see what we get, if we don't get any further?

CO-CHAIR CROOKS: I'm not holding my breathe on them getting the performance data in time. So maybe we should --

DR. PACE: Okay. All right. So we can resume this if need be, okay?

CO-CHAIR CROOKS: Right. I think we're better off, with about an hour left, we should take on one more.

DR. PACE: Okay. Are there any

other, either in the vascular access group, in the patient indication quality of life group or adequacy group of measures that people think would benefit from the full Committee discussion?

DR. KLIGER: Well, I'd love to see one of the quality of life tools. We haven't talked about that before, and I know Andy is just aching to lead the discussion.

DR. PACE: Okay. I think we'll need to review one of the patient education ones. Unfortunately, the quality of life measure group was not able to complete the submission. So we really don't have the testing data. Okay.

And we had sent it, actually, thinking we were going to get some more information. It is something we'd like to have a discussion with you about because it's an extremely important area. The measure that actually got endorsed last year was a process measure of simply using the quality of life

assessment, and there's certainly a lot of interest in actually having a patient reported outcome measure using that data, which is what the preference would be, because, obviously, just collecting that data doesn't necessarily do anything. But, of course, that's another whole measurement issue in itself.

Lauren and I had an initial discussion with Tom Dudley at CMS because we're interested in this measure, a lot of people at NQF, about whether CMS could consider starting to take this on. And, you know, there's certainly some interest, but we have to continue pushing on that. But maybe we'll take a few minutes before we talk about one of the patient education measures to see if any of you have any suggestions or know of people who would be willing to take on a measure of quality of life where it's actually using quality of life data and doing something from the standpoint of patient reported outcome.

Lisa?

DR. LATTS: Well, what I know, and I don't have any answers, is that there's a subcommittee of the QASC that I'm on, the Quality Alliance Steering Committee, a subcommittee called the Patient Reported Measures -- as you know, Karen -- Patient Reported Measures Work Group that is led by Debra Ness and Michael Barr from ACP.

And so we're in the process of going through sort of all the measures that are out there, and I'm not sure if there's something that can be gleaned from that Work Group that would inform this process.

DR. PACE: Right. And I'll just mention NQF is actually starting a project that I'm going to be involved in that's on patient reported outcomes. And we're doing an initial project related to the methodological issues. So I'll just give you a brief -- you know, we've dealt with huge methodological issues for all the measures. And in some ways

they pale in comparison to when we start talking about patient reported outcomes.

So even though these instruments have often been considered very reliable and valid when you're doing patient level measurement and have been used in research studies when you have random assignment of patients to treatment and non-treatment groups, when you start thinking of then taking that data and aggregating it to get a facility level performance measure, you have to think about risk adjustment, you need to think about do you aggregate it at, like, an average, percent improved, percent who achieve a benchmark? There are many big issues with that.

So, that's what that project that's starting up very soon is really going to try to delve into some of those methodological issues of taking these very good reliable and valid patient reported outcome measures at the patient level and what needs to be done,

what's the pathway to getting them to being useful as a performance measure.

Michael, I think the VA has done some work, maybe not on that particular --

DR. FISCHER: My experience with this has been with in the ASC and CRIC cohort studies in chronic kidney disease where we've looked at QoL with SF36 and then the KDQOL in CRIC.

But I think you've outlined very significant methodologic challenges. I mean, it's one thing to assess it, which I think is probably not so controversial, but to move past that and then try to relate that to an outcome measure and somehow, as you said, kind of risk adjust I think will be no small task, which it sounds like you guys are kind of deep in right now already.

On the VA side of things, Karen, I don't know, at least in terms of CKD and ESRD there's a lot of talk about patient self-management and getting data with patient

reported outcomes. But I don't know of formal research, at least that I'm aware of, in the specific domains of CKD and ESRD.

DR. PACE: And maybe I'm going to back up here and say maybe it's worthwhile talking about that quality of life measure. Because I'd like to see -- I mean, if this Steering Committee really feels that it has some value in moving forward, we can pursue more discussions with CMS as being able to collect that information.

I mean, obviously the KDQOL has, from the patient level data, there's reliability and validity information. It's just the process measure has never been implemented, tested. And so I don't want to prematurely cut it off and I'd like to see if you all have any suggestions of a path forward or how you would like to -- and I forget who we had review that. But, go ahead.

CO-CHAIR CROOKS: Yes, Harvey Wells.

DR. PACE: Harvey, yes. You looked

at the measure what was there, so --

MR. WELLS: Yes, I figured Lauren gave this to me because it was doomed to fail.

I do think its important. I remember when I filled this thing out in center and when I filled this out after I was at home, it just struck me my answers were so different. And I think it's important. I mean, as we talk about all these measures, I mean a lot of them are based on lab outcomes and whatever. But I think what's really important to patients is, you know how has it changed their quality of life? Are they able to continue with their lives as they want to or as they choose? And I think to me real true quality measures from a patient perspective is how it's affecting my life. And I can tell you, I mean I've experienced two different outcomes. And the one I was able to continue my life and one I thought my life was over.

So, I do think it's important. You

know, this measure as its presented did not have sufficient data to evaluate it and review it. But I do believe that its something that's worth pursuing and getting the patient perspective on how they feel they're treating someone.

DR. PACE: Right. Connie?

MS. ANDERSON: The KDQOL is also a part of the conditions for coverage and under -- and it's used by the facilities in their quality improvement. And so those patient-related measures within the KDQOL that are below average are what the facility are supposed to be focusing on for quality improvement. And so there may be a way of using that as the percent of patients that fall in that below average category and then showing improvement as you do interventions for the kind of care. So there might be something there that might be able to --

DR. PACE: So if it's mandated, is it mandated that every patient have QAL?

MS. ANDERSON: Every patient except those with these exclusions that are in the denominator exclusion are the same exclusions that are in the conditions for coverage. And the surveyors do review this at each survey, and it's the percent patients that have a below average score and then what they want to see as a plan of care attached to that and how you're going to improve that below average score.

DR. FISCHER: I just think that there is evidence. I mean, I think the importance of assessing QOL and the relationships, at least the epi-relationships between QOL and mortality and other outcomes, there's reasonable evidence in CKD and ESRD, I guess. But moving past that in terms of this has come up with other things: What do you do specifically to improve QOL and where's the evidence for that and if that occurs, does that lead to a change of a outcome downstream or is QOL itself a defined outcome like

mortality? I think those are areas that there's not a lot of evidence I'm aware of.

And you could argue that quality of life doesn't have to be linked to something like mortality or hospitalization. In and of itself could be a defined terminus of an outcome.

DR. PACE: Right.

DR. FISCHER: But even then you're left so QOL because there's a mental health -- there's different composite scores. That's an MCS and a PCS. I mean, then which part are you exactly intervening on and where's the data that that actually changes things? And what would be those processes?

I'm assuming those are the types of things, Karen, that you all may be kind of working through now?

DR. PACE: Well, that is one of the -- I mean, you know we're going to be having some white papers on the methodological issues, but that is one of the questions about

sensitivity to change or clinical intervention, you know doing condition-specific things versus more global patient reported outcomes.

So, Alan?

DR. KLIGER: Yes. I mean there's a basic difference here, though, I think is critical to define. The KDQOL and the other tools we've used, doctors have made up, social workers have made up. We kind of come up with these categories and then validate them and see each of the dimensions. And each study we've done, like we've done at HFM, we've got lots of good data on those objective measures. But the patient-derived measures are just a different realm.

And I keep hearing that our measures, the ones that professional people design, have their place in importance. But we haven't paid nearly enough attention to the patient-defined measures. And to me that's the area that we I think we need to pay more

attention to and then develop ways of examining that here at NQF.

MS. LeBEAU: Not surprisingly, of course, I absolutely agree. I think, you know we talk about this a lot, of course, within the patient advocate community that I work with. And it's functional wellness. It's participation in life. It's all of the things that are very intangible and tough to quantify, but that are very meaningful.

And, yes, with all due respect, of course, the tools that we've come up with so far are useful, but they always tend to have sort of a clinical perspective in there. And this is a little different.

So, I think Alan's point is extremely well taken. Thank you.

DR. NALLY: We happen to be sitting in a room of people that are interested in kidney disease. But this issue really has brought up application to anybody with a chronic medical disease. And I wonder what

NQF's position is more broadly in chronic disease management in patient quality of life issues? I wonder do the heart failure people or any other medical/surgical specialty seem to have an inside track on getting their arms around this issue where we might learn from them, or are they in the same kind of dire straits we are?

DR. PACE: Well, I can tell you that I think everyone's kind of at the same place. There have been things brought in to other projects, and I know in the cardiovascular project, for example, one of the -- you know, if it was the Seattle Angina Questionnaire or some patient reported measure, but the issues about what's the performance measure. You know, everybody agrees that's a reliable and valid measure at the patient level, but what are you suggesting we do at the performance measure level?

I think the only one that I can mention right off that has NQF endorsement,

and it may gotten it as time limited, was bringing in a depression scale, patient-reported depression scale. I believe it was the PHQ9, and having a performance measure based on change, I think. And I don't have the details about it.

But in terms of these issues it's really across the board that people are struggling with. And that's one of the reasons we're doing this project to look at the methodological issues more across the board, because there's a huge clamor for performance measures based on patient-reported data and the things that matter most to patients; function, well-being, those kinds of things.

And even from the standpoint of, I know from the eye surgery group, you know they're looking at patient-reported visual function after eye surgery, which you know that's what matters. Does the patient think they can see? And people are looking at those

in terms of after knee and hip surgery. But this bringing it to the level of a performance measure has been -- it's not solved anywhere that I know of.

DR. KLEINPETER: So, Karen, one other question. What about the ambulatory care project. Because I remember some years ago when I was on that project that there were some things for depression and anxiety. Did those -- one of them was time limited, but I think the other one didn't pass. Did they have any --

DR. PACE: Was it an actual patient-reported scale?

DR. KLEINPETER: It was patient --

DR. PACE: I can't answer that.

DR. KLEINPETER: Okay.

DR. PACE: I'd have to check.

I mean, the other thing as you all know and I should mention, too, NQF has endorsed the measures associated with the CAPPS instruments. And in the last project

the ESRD CAPPS was endorsed. And its due for endorsement maintenance. And the reason you don't have it in your materials here is because AHRQ has had some cutbacks and they didn't have the resources to maintain the measure in time for this project.

Again, we've had some conversations with CMS about that because CMS was very interested. And CMS and AHRQ are now talking about maintaining that measure. And, luckily, NQF is going to be doing a project I think early next year specifically on patient experience. So we'll be able to -- that measure will continue to be endorsed and it will come through endorsement maintenance with some other patient experience measures. So I just wanted to kind of assure you that's not going away, but it's kind of the realities of resources at this point in time.

Okay. So maybe what we can do is at least begin going through one of the patient education measures. They're similar;

one's facility and one's physician level. And then we'll probably only get through one of them, but I think then it'll be easy for us to pick up on the other ones. So --

CO-CHAIR CROOKS: We should stop at 3:00 so we have time for comments.

DR. PACE: Yes.

CO-CHAIR CROOKS: Next steps and adjournment by 3:15.

DR. PACE: Right. Okay.

CO-CHAIR CROOKS: Okay.

DR. PACE: So let's do the facility level one. 0324.

MS. McGONIGAL: Karen, do you want us to start with remarks?

DR. PACE: Oh, I'm sorry, yes. Yes. So, Lisa, do you want to present the measures?

MS. McGONIGAL: Okay. Again, both of these measures are from the Kidney Care Quality Alliance. We've submitted measure 0324 Patient Education Awareness - Facility

Level and 0320 Patient Education Awareness - Clinician Level. Those measures were endorsed by NQF in 2008 and are included among CMS' Phase III clinical performance measures. The Phase III CPMs are slated for us by CMS in its CROWNWeb dialysis facility data repository when it becomes functional.

The physician level measure was field tested in clinician officers, coincident with the AMA PCPI Renal measures and the facility level measure was tested at 53 dialysis facilities across the United States.

The underlying rationale for both measures, which are identical as Karen mentioned except for the level of analysis, is to ensure that all ESRD patients are educated on all available renal replacement therapy options: Hemodialysis, home hemo, peritoneal dialysis, transplants and identification of living donors and no or cessation of renal replacement therapy at least once yearly.

The measures are consistent with

the CMS conditions for coverage and a body of evidence demonstrating that patients knowledgeable about dialysis are more likely to use a AVF as vascular access, have less depression and improved medication adherence and treatment attendance. And are more likely to survive and to get a transplant than their less well informed counterparts.

In particular, we'd like to reference a June 2011 study that wasn't included in the initial measure submission form because its too new. The study demonstrated that attendees of the National Predialysis Treatment Program that provided education about modality options more frequently selected home dialysis and had lower catheter rates and mortality risks during the first 90 days of dialysis when compared with period prevalent incident patients who didn't participate in the program.

In the study the unadjusted early

mortality hazard ratio is found to be 0.51 for program attendees and after adjusting for case mix and laboratory values, the hazard ratio was 0.61 per program attendees. In all outcomes, P was less than 0.001.

Also, I'd like to note an error that was in the measure submission form regarding the clinician level measure. Under "Summary of Evidence For Performance Gaps," which is section 1B.2, the form indicates that the performance rate in physician's offices during field testing was 97 percent. What should be indicated is that the rate when assessing the number of patients educated on at least one renal replacement therapy option was 97 percent.

An additional paragraph was omitted in which it was noted that to receive credit for the measure patients must be educated on all six of the modalities addressed in the measure and none of the patients included the sample methods criterion said that the

physician level performance was actually zero percent.

The facility performance rate, as we accurately noted in the measure submission, was 16.4 percent during field testing, meaning that there was a significant gap in care in both settings.

And we would again like to thank you for your consideration of the measure. And we welcome any questions now or after your deliberations.

DR. PACE: And actually, I can let you guys decide, Andy and Kathy, which measure you want to talk about or if we can talk about them today?

DR. NARVA: It's the same measure.

DR. PACE: It's the same measure. And if there are issues, we can bring them up. Okay. So Kathy, do you want to start?

DR. NALLY: Before you start.

DR. PACE: Yes.

DR. NALLY: Is it possible to ask

them one specific question about the information that could not be presented because of the newness of the information?

DR. PACE: Yes.

DR. NALLY: Clearly, earlier in the equation we could have the patient educated and give them options, perhaps the better for everyone involved. How was it that those patients were identified and able to participate in a pre-ESRD study?

MS. McGONIGAL: Okay. This is the Laxson, et.al. paper that was published in June in the American Journal of Kidney Disease. It was done at Fresenius Medical Care. I don't have the exact how they were able to identify the patients, but they were all within Fresenius, so they were recruited that way. Similar to what they did for their Right Start Program when they studied that.

Does that answer your question?

DR. NARVA: Actually, the Right Start data that you cited cites incident

dialysis patients. Yes. And so is it the same curriculum but a different group of patients?

MS. McGONIGAL: Yes, this is a different curriculum. They focused specifically on educating the patients on available modality options rather then going into all of the stuff that the Right Start did. It focused just on just TOPS. Yes.

DR. LATTS: Excuse me. Can I say Right Start is different from TOPS? Yes. Okay. I'm sorry.

DR. PACE: Kathy, do you want to give us a description of the measure and then we'll get into the rest.

MS. LeBEAU: Yes. Thank you.

Well, we are looking at these two very similar measures. It is a percentage of the physicians end stage renal disease patients aged 18 years and older with medical record documentation of a discussion of renal replacement therapy modalities to include: Hemodialysis, peritoneal dialysis, home

hemodialysis, transplant and identification of potential living donors as well as a no treatment order or cessation of treatment option at least once during the 12 month reporting period.

The numerator would be the number of patients from the denominator, again with medical record documentation, that a discussion did occur including all of those above listed options. And the denominator would be all of the ESRD patients aged 18 years and older.

Feel free to step in, Andy, at anytime.

Talking about impact, high impact, education programs for chronic kidney disease patients have shown to delay the time onto dialysis and improve survival. And it indicates that patients with greater knowledge about dialysis at initiation are more likely to use an AV fistula or graft than a catheter.

The Right Start patients that we

were talking about have significantly improved mental composite scores and reduced hospitalization and mortality rates compared to control subjects demonstrating that such a structured program of prompt medical and educational strategies in incident hemodialysis patients resulted in improved morbidity and mortality that lasts up to a year.

DR. NARVA: Well, you know since a third of our patients meet the nephrologist when they're having a catheter inserted, it's not hard to argue that there's an educational gap, you know.

I think a lot of the data that's presented concerns pre-dialysis; education and its impact prior to initiation.

And I think overall one of the issues in looking at these two measures is clearly there's a big educational gap, whether this measure would address that educational gap.

CO-CHAIR CROOKS: The horse is out of the barn, in a sense? Because the denominator is ESRD patients on dialysis.

DR. NARVA: Right. And, you know most of what's cited and most of the experience relates to interventions that are done prior to initiation of dialysis.

CO-CHAIR CROOKS: Right.

DR. NARVA: There's very little to support the kind of intervention that's described in this measure.

CO-CHAIR CROOKS: A related issue which may be better -- I'm not sure this comes under validity, but this is really just looking for check marks, in a sense. You know, there's a note in the chart. Does that equally effective education? I'm not sure where that should be discussed or considered.

DR. PACE: Probably under validity.

So, Connie?

MS. ANDERSON: Just another comment about this is it's also participation and the

conditions for coverage issue as well, and it is that facilities are required under the conditions of coverage to provide modality education in all of these topics. I think it's within the first six treatments and then yearly thereafter. And there's not a measure of the quality of the education, it's as you said Peter, it's a check box that the patients have been educated on this.

CO-CHAIR CROOKS: Yes.

MS. ANDERSON: So this is also a measure that's being monitored through CMS through the survey process.

MS. LeBEAU: It is. And while you're right about the not addressing the quality of the education, they do specifically say that whether or not the facility offers the treatments, they have to educate on them. Which I think, frankly from a patient's perspective, has been historically a problem. So there is that particular stipulation.

DR. PACE: So maybe what we'll do

is -- I mean, obviously you have some questions about the measure specifications. So I guess first let's try to go back to impact. And I guess the question does patient education impact outcomes. And I think you're right, then the question is: Does this measure actually fit with the opportunity for improvement and evidence, et cetera? Does that make sense to everyone on the Committee?

CO-CHAIR CROOKS: Well, whether or not this effectively causes changes in outcomes, I think it is important that it should have high impact.

MR. McMURRAY: Just a clarification. The Right Start Program and the impact programs both are not predialysis, they're both in the first 90 days of dialysis. So it is on folks who have already started.

MS. LeBEAU: Well, this does define the -- excuse me. The numerator as ESRD patients. But certainly there's no argument that CKD patients probably need it even more.

CO-CHAIR CROOKS: So I think unless someone has a burning issue, we can at least vote on the impact: High, moderate, low or insufficient. Are we ready? All right. Let's go.

MS. RICHIE: Lorien, you still there? Impact?

DR. DALRYMPLE: For impact moderate.

CO-CHAIR CROOKS: There's 21. So we have 11 voting high, nine moderate and one low.

Okay. Now onto the performance gap. And just as long as my mic's on, this is a required Medicare condition for coverage. Can we assume it's always being done, and therefore there's no performance gap? I mean, you don't get paid without it.

DR. PACE: But the data presented--

CO-CHAIR CROOKS: That was just a--

DR. PACE: You guys, Andy and Kathy --

CO-CHAIR CROOKS: Prove me wrong.

MS. LeBEAU: One would assume that, but according to the conclusions from the studies that are cited in this, the findings are that at both the facility the physician's office level indicate that a majority of ESRD patients are not being educated on all renal replacement therapy options. And also, that provider performance varies significantly by modality, again leaving out treatments that they may not offer. So it did identify a significant medical gap.

CO-CHAIR CROOKS: So this is based on looking for documentation as opposed to asking the patient whether they received education, is that right? Okay.

DR. FISCHER: So it was a gap then maybe in documentation, not actual --

DR. NARVA: Maybe there's a gap in education, but no gap in documentation.

The USRDS when they did the -- they reported on data for Meeting Healthy People

2010, they reported data on percentage of patients who had a discussion of transplant. And even though it was very high, but you know I think that that's a box. Is that a box on 27 or 28, or somewhere along the way. So I think the point that Karen raises is very important. It's one thing to have a sort of a check-off box. It's another thing to have some documentation and some patient understanding

DR. WELCH: Well, and it's not just understanding. It's effective decision making.

DR. NARVA: Sure.

DR. WELCH: So there's a big leap here about --

DR. NARVA: The self-management.

DR. WELCH: -- I've done my job. I've given you information and then what happens to that information? We are making a leap.

DR. BERNS: Just a question about

the performance gap. Is the assessment done after or sufficiently long after this had become a condition of coverage?

MS. LeBEAU: I'm sorry. Before.

DR. BERNS: So it really isn't evidence of a current performance gap?

MS. LeBEAU: Could you please clarify? I'm sorry.

DR. BERNS: My suspicion was, which has proven to be correct, is that the assessment of the performance gap was prior to this becoming a condition of coverage. So that since its become a condition of coverage, we don't have evidence of a performance gap.

CO-CHAIR CROOKS: Okay. More? Yes?

MS. WAGER: Excuse me. Can I make a comment to Dr. Narva? Sometimes patients are sent for education maybe a year out before they need dialysis. So they've been educated. Some of them have a fistula, some of them may not. And they come to the clinic and they get

-- they're assessed, and then they're assessed did you attend the TOPS class, were you educated?

Well, I remember when I was on dialysis. I forgot a lot of stuff. You know, so the gap could also be that the patient doesn't remember. Because we do have some patients, I had one patient that she came to a class four years before she started dialysis. So --

CO-CHAIR CROOKS: Well, but that's why I asked the question, too, of is this performance gap data based on documentation rather than asking the patients what they remember. And I was told, yes, it is.

MS. ANDERSON: No, it's not.

CO-CHAIR CROOKS: No, it's not?

MS. ANDERSON: It's not.

CO-CHAIR CROOKS: I'm sorry. Well, please explain some of it.

MS. ANDERSON: It's based on at the point of time within the first six treatments

that you are obligated to educate the patient on each of these conditions. So each of the treatment modality options. And what your documentation is is that, yes, you have educated the patient on each of those. And then --

MS. LeBEAU: That's looking forward to provision and conditions --

MS. ANDERSON: That's the way the conditions for coverage are written, yes.

DR. VELEZ: That's not this measure. Yes, this measure is only documentation that this happened, whether it was ten years ago or two days ago --

DR. NISHIMI: No. It's documentation within the year.

DR. VELEZ: In a 12 month period the documentation.

DR. NISHIMI: Right.

DR. VELEZ: The documentation could have happened at the office level.

MS. LeBEAU: But I do think the

salient point from what Bobbie said is that exactly the percent that Dr. Narva cited, a good third of these patients are being educated at a time when they are overwhelmed with a new diagnosis. They're sick. They're starting dialysis treatment. It's not a great time to do education. So, I think that's the very important part about it having the 12 month and repeated.

Also things change. You go from one modality, you are transplanted, you go back to dialysis. Very important that that opportunity be repeated.

DR. VELEZ: Again, the way I read this measure is documentation that this was explained. Again, this could have been done a year before and there's documentation in my chart today that I did this last year. And that's all that it requires in that 12 month period. That's the way I read this measure.

MS. LeBEAU: No, it's --

MS. ANDERSON: You're correct, but

within the conditions for coverage you're obligated to repeat it. Yes. And I think the performance of the -- gap performance is based on pre-condition for coverage patient education.

DR. PACE: But the specifications say at least, and we'll ask the developer. The specifications say at least once during the 12 month period.

MS. McGONIGAL: Right. If the education occurred at least once during the 12 month period. Documentation that the education occurred at least once per year.

(Simultaneous speaking.)

DR. PACE: Okay. So let me ask it this way, because I think this is your question: So you made document it every year, but your documentation may be that I told them two years and I --

MS. McGONIGAL: No. Documentation that the education occurred at least once a year.

DR. PACE: Okay. All right. Got it.

CO-CHAIR CROOKS: Okay. Thank you.

DR. WELCH: So it doesn't mean that they heard it, is that what I'm hearing?

CO-CHAIR CROOKS: Well, we understand that.

DR. WELCH: Okay.

CO-CHAIR CROOKS: But in terms of trying to judge the performance gap, we need to know that this metric was done and the data that we have here is that depending which modality you're talking about, the gap was -- the performance was between 30 and 80 percent, depending on the modality. Am I reading that right? Okay. So I judge that to mean there is a performance gap, so that's what I'm going to vote. And are the rest of you ready to vote? Okay. High, moderate, low or insufficient.

MS. RICHIE: Lorien?

DR. DALRYMPLE: For performance

gap, high.

CO-CHAIR CROOKS: Four votes high, 10 moderate, one low, six insufficient. Okay.

You must have voted insufficient.

Okay. So we're to the point where we can -- this is a process, not a health outcome. So we can look at the body of evidence.

DR. PACE: Right.

CO-CHAIR CROOKS: Andrew or Kathleen, somebody want to step us through it quickly?

DR. NARVA: This from the application and this focuses on renal replacement modalities, which says "While several studies have demonstrated an association between patient education and improved outcomes in the ESRD population, none were identified that focused exclusively on renal replacement modality options as is the case with this patient education measure."

CO-CHAIR CROOKS: So the quantity

is zero or it's not closely related to the metric?

DR. NARVA: The evidence out there doesn't relate to this measure.

CO-CHAIR CROOKS: The evidence says that education leads to better outcomes, kind of a general --

DR. NARVA: Yes.

CO-CHAIR CROOKS: -- in all settings or pre-dialysis settings?

MS. McGONIGAL: Yes. We asked you to consider the supplemental study that we've included since then, the TOPS study as well. And that's the only one available at this point in time on pre-dialysis modality education.

DR. NARVA: But that invention is also very different from -- that's an extensive curriculum, is that correct?

DR. PACE: So let me just kind of bring us back on evidence. You know, obviously it would be indirect evidence and

require some assumption.

The other thing is that we do if you wish to invoke it, we do have an exception for areas where there's really not going to be evidence and it's based on expert opinion.

So we could rate this body of evidence on patient education that would be indirect, which is part of the quality assessment. And then we can talk about, you know if the evidence is really not sufficient, then the next step would be whether you want to move forward based on expert opinion. Does that make sense?

CO-CHAIR CROOKS: So could we move to agree that the body of evidence would not be sufficient but that -- okay. I was going to try and save a couple of minutes. Okay.

So let's vote on the quantity.

DR. PACE: Okay. Go ahead.

CO-CHAIR CROOKS: Okay. So one high, two moderate, three low, four insufficient.

MS. RICHIE: Lorien, quantity?

DR. DALRYMPLE: Low.

CO-CHAIR CROOKS: So we have two votes moderate, six low, 13 insufficient.

Okay. Quality of body of evidence, shall we vote? Okay. Turn on the clock. Thank you.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Insufficient.

CO-CHAIR CROOKS: Okay. And the results are one high, three moderate, four low and 13 insufficient evidence.

And consistency?

MS. RICHIE: Lorien, consistency?

DR. DALRYMPLE: Insufficient.

CO-CHAIR CROOKS: Because there's insufficient evidence, there's insufficient consistency. Okay. Eighteen insufficient, four low, one moderate.

DR. PACE: Sixteen.

CO-CHAIR CROOKS: Sixteen -- let's try that again. Sixteen insufficient, four

voted low, one voted moderate.

So now we can get to the point where we may consider overriding this due to expert opinion?

DR. PACE: Right. Right. So next slide.

CO-CHAIR CROOKS: If there's no empirical evidence and expert opinion is systematically assessed with agreement that the benefits to patients greatly outweigh potential harm, is it judged that potential benefits to patients clearly outweigh potential harms? Can we just go ahead and vote?

DR. PACE: You guys ready to vote or you want to discuss?

CO-CHAIR CROOKS: Did I state it clearly? Okay. Let's vote.

MS. RICHIE: Lorien, yes or no?

DR. DALRYMPLE: Yes.

CO-CHAIR CROOKS: Okay. It is a considered opinion of this august body that

the expert opinion should carry this measure forward; 18 yes, three no.

DR. PACE: Okay. So I think then we passed importance to measure and report. Yes. Okay.

So I know --

MS. LeBEAU: You're pushing the envelope. We have ten minutes.

DR. PACE: Okay. All right.

CO-CHAIR CROOKS: We can do this in ten minutes.

DR. PACE: Okay. Good.

CO-CHAIR CROOKS: All right. So reliability testing. This is an existing is an existing metric, right?

DR. PACE: Right.

CO-CHAIR CROOKS: So there should be some data on --

DR. PACE: Right, and there is.

DR. NARVA: The Right Start that was cited, I think only 16 percent of patients were educated on all modalities.

DR. PACE: Okay. And what we're going to look at now is the specifications and the reliability testing for this measure. So under 2.A.2 they did some testing in both the facilities and physician office. So they did inter-rater reliability and provided data on that. And I don't know, Andrew, you want to say anything about that? I'm trying to see if I can pull up the --

DR. NARVA: I think the issues there related to defining what education was.

CO-CHAIR CROOKS: A kappa statistic of .0026 for inter-rater reliability looking at the same data being extracted by two people, right?

DR. PACE: Yes.

CO-CHAIR CROOKS: Is that a low kappa?

DR. PACE: What was it?

CO-CHAIR CROOKS: .0026. With a 95 percent confidence interval.

DR. PACE: Is this in a table or --

CO-CHAIR CROOKS: I'm looking at it here.

DR. PACE: Yes.

DR. NISHIMI: We want to note that we're talking about the facility measure, right?

DR. PACE: Yes.

CO-CHAIR CROOKS: Yes.

DR. NISHIMI: Because the reliability statistics differ.

MS. McGONIGAL: Table 2.

DR. NISHIMI: Table 2 Attachment A.

DR. PACE: Okay. So we need to open up the --

DR. FISCHER: Yes, I think there's a decimal point error in that kappa.

MS. McGONIGAL: That is correct. It's negative 0.0026.

DR. FISCHER: Oh, that's a negative?

MS. McGONIGAL: Yes.

DR. PACE: So do you want to comment on that Lisa?

DR. NISHIMI: This is why we don't think that it can be done in 10 minutes.

MS. McGONIGAL: Yes. Right. Yes.

So based on the literature, negative kappa value indicates that the auditor obtained the same results as the facility abstractor, less then would be expected by chance alone.

There was also relatively low concordance rate, again demonstrating substantial interabstractor disagreement. However, when we reviewed this data we did not believe that the negative kappa and low inter-rater concordance was due to unreliability of the measure specifications or tool, per se. Because the type of error was not random and all of this is demonstrated in the tables here. Rather significantly more errors were missed information that led to underreporting, in other words false negatives. So when we went back into the facilities to review the charts, they had educated on various things

that they had not given themselves credit for.

Further, the underreporting often stemmed from an apparent lack of understanding by some facilities as to what constituted education and was documented in the records for the purpose of the measure specification.

One particular problem was end of life discussion and advanced directives regarding cessation of renal therapy.

Other facilities seemed to get it and did perform very well. So we just thought that it was, perhaps, that some facilities were not educated well enough on how to collect this data.

Distribution around the facilities. The errors among the facilities was not even. There was a bimodal distribution, again suggesting that some facilities got it and some did not.

And when we went into the physician's office there was almost perfect reliability between the two expert

abstractors, the people who knew what to look for and they were able to get a very high kappa of 0.8474.

So, we performed some additional facility-by-facility error analyses and reliability analyses by data element. And these are also described in detail on the major submission form. We believe that it demonstrates that the patient education measures can be reliably collected and that the negative kappa for the overall patient education measure performance is not an indication that the specifications are unreliable.

We believe that improving the instructions and educating facilities to recognize what constitutes meeting the specification should reduce the high numbers of false negatives. Again, when reduction scenarios of the high false positive rates were analyzed, kappas indicate excellent agreement and reliability.

Also, ongoing implementation of the new conditions for coverage which require these education modalities be discussed, we believe it will improve the reliability by sensitizing facility personnel to organize their record keeping better so they will be more able to reliably collect the data element.

We also wanted to note that when we were going in over the course of the year of data collection, we noticed that the facility's way of keeping track of this was actually changing over the year as they were becoming use to the idea of conditions for coverage. So they were already improvising and coming up with new ways to track this data.

Finally, implementation of CROWNWeb and accountability for patient education can improve reliability by deploying more detailed instructions and training, and by sensitizing facility personnel.

So that is --

DR. PACE: So I think that -- because it's the same data that you collected looking in facility records and physician records. And the difference was you had two kind of expert abstractors versus a facility person and an expert abstractor?

MS. McGONIGAL: That's correct.

DR. PACE: Okay.

CO-CHAIR CROOKS: Yes?

MS. ANDERSON: I'd like to ask the developer, right now these patient education measures are not a part of CROWNWeb. And at this point, at least having been active in the CROWNWeb process, I don't know that they are going to part of the CROWNWeb.

DR. NISHIMI: All we can do is report that we had a conversation with CMS last month and they remained very interested in pursuing this as an incorporation. But the time frame for that for that build out, is obviously something we don't know.

DR. PACE: Other questions or discussion about reliability? So I think what their data shows is that there's the potential to have a reliable measure, and most of the testing we get is on a small sample and shows a potential. I think you have to weigh the difference in the methods and in terms of looking at these results.

CO-CHAIR CROOKS: So we're not going to get through this measure, apparently. So should we go ahead and vote on reliability or would people like to think about it a little bit more?

I see we're getting some tokens held up in the air, spinning around in circles.

DR. PACE: Okay. Well, why don't we vote on reliability and then we can pick up this measure later.

CO-CHAIR CROOKS: Okay.

DR. PACE: Resume it at our first opportunity.

CO-CHAIR CROOKS: Okay. So let's vote on reliability: High, moderate, low or insufficient evidence.

MS. RICHIE: Lorien?

DR. DALRYMPLE: Moderate.

CO-CHAIR CROOKS: And the final vote of the day, 11 moderate, eight low, two insufficient. So if you add insufficient to low, moderate barely carries. Eleven to ten.

So 11 moderate, eight low, two insufficient. Thank you.

So we're at that point where we're going to stop our evaluation metrics. We will, first of all, open the phones and the floor for public comment. So does anybody here or on the phone wish to make any more comments at this time?

Okay. Well, that's --

DR. PACE: And we have some audience, too.

CO-CHAIR CROOKS: Yes. Measure developers, anybody else in the room, on the

phone? Okay. Thank you.

So, Karen, how are we going to proceed from here? Have you and Lauren got it all figured out now?

DR. PACE: The first thing is scheduling conference calls. So you will be getting emails from us from us very quickly to get some calls set up. And we'll be working on a process to try to accomplish the rest of the measures.

I think it helped that we had some discussion in all of the topic areas, because I think that will ground us going forward. So I appreciate that.

Jeff?

DR. BERNS: Given what I'm sure is going to be great difficulty in getting the conference call with this group, would it be possible or would it make sense to divide into two or three groups and try to get the work done that way based upon just availability. So if you a third or a half of people

available for one call and you do it and another after another.

DR. PACE: Yes, we can certainly look at all those options. And --

CO-CHAIR CROOKS: But we do need a confirming vote.

DR. PACE: Right.

CO-CHAIR CROOKS: And we can do all the voting into the computer system.

Although I have to say, Karen, when I wanted to get a metric to come back up again, putting my name in and putting the same number and it gave me a clean sheet. So if I don't like the way I voted before, am I stuck with what I did.

DR. PACE: No, no. We would have to sit up a different tool for this.

CO-CHAIR CROOKS: Okay.

DR. PACE: So that you could go back. So we have a lot of kind of logistical things to try to think out how to best move forward and coordinate with your time. And

you know, be most efficient and thorough.

So, you know if you have some suggestions, you know I think certainly if we need -- we don't expect that we'll ever get a 100 percent on a conference call. But we'll, you know we'll generally look at multiple options and pick the option with the most. But we may have to do several calls and we'll have to move forward with a substantial majority versus 100 percent. We won't --

CO-CHAIR CROOKS: So let me kind of summarize some next steps a little more concretely.

It'll be expected that the Steering Committee members will at some point in time, and they can't start right away because if you go home tonight and start putting in votes, they're not going to count.

DR. PACE: Yes.

CO-CHAIR CROOKS: But at some point in time you'll be instructed to finish your evaluation of the measures and to vote. And is

that--

DR. PACE: Right. So let me ask you this, because it was kind of where I was going originally this morning.

We have two ways we could do this. One is to get together on a conference call and have more discussion, and then vote. The other way would be to set up a voting on the measures that we have yet to vote on. Invite everyone to do that before the call and then use the call to review those results and discuss any discrepancies or potential areas where there were issues.

So I want to just get a feel. I mean, these are --

CO-CHAIR CROOKS: Well, one difference between what we were proposing this morning and the situation we're in now is that --

DR. PACE: We were going to have some discussion.

CO-CHAIR CROOKS: -- we were going

to have discussion, right? And if we just go back and start voting, we won't have had an opportunity for discussion. And we need to hear Alan's opinion or we can't vote intelligently. I mean, let's face it.

DR. PACE: Right.

CO-CHAIR CROOKS: As well as many other people.

So maybe another option, this is where smaller groups could come in, too. For instance -- I'm just thinking out loud, but let's say a group of mineral enthusiasts got together and they discussed and voted, what would we do with that? Would that help us? Or we still need to come back --

DR. PACE: Yes, I think we still need to come back. Yes.

DR. LATTS: I would suggest that you set up calls by domain and use a Doodle survey to set up the calls. You set up the time where the measure reviewers all agree they can attend with the rest of us optional

as schedules allow.

The measure reviewers review the measures, you know come up with their votes on each thing. We as a -- then we as a group come together and then can just quickly go through based on that.

DR. PACE: All right. So we'll, like I said, we have to go back and think about logistics and maintaining the integrity of the process. And we'll get with you as quickly as we can, but we will start getting schedules as quickly as possible.

CO-CHAIR CROOKS: So don't start voting on anything yet until you get instructions. But please be looking for and respond to meeting invitations as soon as possible. We want to get that calendared as soon as possible.

DR. DALRYMPLE: Karen, is it possible to have the stewards present at the time of final voting, if at all possible? Because I think it really helps with some of

the clarification and --

DR. PACE: Yes, definitely. All the conference calls will be open and stewards invited and open to the public. Yes, definitely.

CO-CHAIR CROOKS: Joe, you were asking what kind of timeline or time frame? Originally we wanted to have the Committee's work done by next week?

DR. PACE: Yes.

CO-CHAIR CROOKS: Last week? So -

DR. PACE: We're just going to have to deal with that. So --

CO-CHAIR CROOKS: To be determined. Okay. So --

DR. PACE: We have reality in our face, so we'll just have to deal.

CO-CHAIR CROOKS: Any other -- at this point we have a couple of minutes left. Would anybody on the Committee like to make any comments about their experience, the process, suggestions for improvement? Myra.

DR. KLEINPETER: One suggestion, in terms of some of the introductory stuff that we went through, perhaps that should be a teleconference a week before the meeting and perhaps having the individual work groups have a one hour call to go over things. That would kind of speed things up so that when everybody's in a group, we may move a little bit faster.

CO-CHAIR CROOKS: Good. Thank you.

Other comments, suggestions?

DR. PACE: Feel free to send us emails and we appreciate all of you.

CO-CHAIR CROOKS: We really, really appreciate your time and focus.

DR. PACE: Thinking power, I know it made everyone tired and we appreciate all the energy and time you've committed. Thank you.

CO-CHAIR CROOKS: Thank you.

(Whereupon, the above-entitled matter went off the record at 3:06 p.m.)