NATIONAL QUALITY FORUM

+ + + + +

RENAL STANDING COMMITTEE

+ + + + +

THURSDAY
MAY 7, 2015

+ + + + +

The Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:00 a.m., Constance Anderson and Peter Crooks, Co-Chairs, presiding.

PRESENT:

CONSTANCE ANDERSON, BSN, MBA, Northwest Kidney Centers (Co-Chair)
PETER CROOKS, MD, Kaiser Permanente (Co-Chair)
ISHIR BHAN, MD, MPH, Partners Healthcare, Massachusetts General Hospital
LORIEN DALRYMPLE, MD, MPH, University of California Davis
ELIZABETH EVANS, DNP, American Nurses Association
MICHAEL FISCHER, MD, MSPH, Department of Veterans Affairs
STUART GREENSTEIN, MD, Montefiore Medical Center
DEBRA HAIN, PhD, APRN, ANP-BC, GNP-BC, FAANP, American Nephrology Nurses' Association
LORI HARTWELL, Renal Support Network
FREDERICK KASKEL, MD, PhD, Children's Hospital at Montefiore
MYRA KLEINPETER, MD, MPH, Tulane University School of Medicine
ALAN KLIGER, MD, Yale University School of Medicine
MAHESH KRISHNAN, MD, MPH, MBA, FASN, DaVita Healthcare Partners, Inc.

LISA LATTS, MD, MSPH, MBA, FACP, LML Health Solutions and CMO, University of California Health Plan
KARILYNNE LENNING, MHA, LBSW, Telligen
FRANKLIN MADDUX, MD, FACP, Fresenius Medical Care North America
ANDREW NARVA, MD, FACP, FASN, National Institute of Diabetes and Digestive Kidney Diseases - National Institutes of Health
JESSIE PAVLINAC, MS, RD, CSR, LD, Oregon Health & Science University
MICHAEL SOMERS, MD, American Society of Pediatric Nephrology, Harvard Medical School, Boston Children's Hospital
DODIE STEIN, PhD, MSW, LCSW, Indiana University Health Home Dialysis
BOBBI WAGER, MSN, RN, American Association of Kidney Patients

JOHN WAGNER, MD, MBA, Kings County Hospital Center

JOSHUA ZARITSKY, MD, Nemours/Alfred I. duPont Hospital for Children

NQF STAFF:

POONAM BAL, Project Manager

ALEXANDRA OGUNGBEMI, Project Analyst

KATHRYN STREETER, Senior Project Manager

ALSO PRESENT:

JOEL ANDRESS, PhD, Centers for Medicare and Medicaid Services
AMY BECKRICH, Renal Physician=s Association

CLAUDIA DAHLERUS, PhD, MA, University of Michigan

BARBARA FIVUSH, MD, Johns Hopkins University*

EDWARD JONES, MD, Delaware Valley Nephrology and Hypertension

JOE MESSANA, MD, University of Michigan

LISA McGONIGAL, MD, MPH, QMRI

PAUL PALEVSKY, MD, University of Pittsburgh*

SARAH SAMSPEL, NQF Consultant

DOUG SCHAUBEL, MD, University of Michigan

DALE SINGER, MHA, Renal Physicians Association

*** present by teleconference**

TABLE OF CONTENTS

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

(8:02 a.m.)

CO-CHAIR ANDERSON: All right. Well, thanks, everyone, for coming a half hour early, so we can get started. We have a large agenda to get through today. Yesterday -- thank you, everyone -- we did make a lot of progress, but we've got a lot more to do.

Just a reminder that we are going to try and get through all of the measures. So if we could keep our comments and not repeat the same comments, that would be very helpful to move through the agenda.

At the end we will have a wrap-up, and we would really like to hear your comments about the process and how things went.

Peter, do you have anything else to say?

CO-CHAIR CROOKS: I'm still -- I'm up first, so I have to get ready.

CO-CHAIR ANDERSON: Oh, you're up first. Okay.

CO-CHAIR CROOKS: I'd just like to add, again, thank you for coming in early. For those of you who enjoyed the swank Helix Hotel, and happened to get charged for the full price, the staff is looking into it, and we'll let you know later if they were able to reverse charges or you'll be billing it back to NQF.

CO-CHAIR ANDERSON: Okay.

CO-CHAIR CROOKS: So we will have more on that later.

CO-CHAIR ANDERSON: All right. Our first measure is 0249, and the developers, Claudia and Joel? You guys are already here. You're right on top of things.

DR. DAHLERUS: Okay. Good morning. So I will keep the opening statement very, very brief, just because some of the things that we discussed with the peritoneal dialysis adequacy measures apply to the HD measure.

So the first one and the prior one being that, as you are aware, this also included the upper threshold of 5.0 for the individual

measure. And as we discussed yesterday, that will be removed. We will redo the reliability and validity testing for this measure and then resubmit updated documentation. And, again, that is something that is very manageable.

The only other thing that I wanted to note, because I think this may have come up in some of the remarks about the measure, is the 2006 guidelines that were cited in our evidence summary, and recognizing that, you know, those are obviously quite a few years old.

However, the HD adequacy workgroup for KDOQI did meet last year, and there were draft guideline updates that were released, and the important point being really no new updates in terms of a revision to the minimum Kt/V for dialysis adequacy.

So I will just I think end my comments there in the interest of time, because we also have similar adequacy measures to move through, and I think a lot of the same issues apply.

Thank you.

CO-CHAIR ANDERSON: All right. And, Peter and Ishir, for discussion of the measure?

CO-CHAIR CROOKS: Yes. Would you like to take the lead?

DR. BHAN: Sure.

CO-CHAIR CROOKS: Yes, please do.

DR. BHAN: I can kick it off? So this is somewhat similar to the last measure that we looked at, in that it is looking at hemodialysis adequacy. It is an outcome measure using claims and electronic data. But this one is at the facility level as opposed to the individual clinician level.

So just going to the evidence, there is -- I think we went through this with the last measure as well. There is the -- as was just mentioned, there is the KDOQI guidelines from 2006, which had a grade A but are somewhat old. There are a number of studies showing clearance correlations with outcomes, and there is the hemo study showing that higher clearances aren't necessarily helpful, at least overall.

So I think overall the feeling was that, you know, 1.2 is a reasonable threshold. That is sort of where things have settled out, and we thought the evidence was, you know, as good as it is for this kind of thing. There is, you know, obviously ongoing debates about optimal clearances, but it's as reasonable as it was for the previous measure we discussed. And we talked about the birth threshold, but I think that has been addressed.

CO-CHAIR CROOKS: Yes. I felt the evidence review was based on KDOQI. Well, to be a little picky, I noticed little things -- and this gets to the issue of filling in the forms and answering the questions, that the answers to 1A 7.7 and 1A 7.8 were identical. They were just cut and pasted in.

And the second one was about harms, and it really didn't address the question. Again, I'm being picky, but I'd like to see the developers, you know, help us out so we don't have -- you know, to take us down the path to see

what it is they are trying to show us.

But the evidence is grade A. There was no new evidence presented since 2006.

So the floor is open for other comments on the evidence.

MS. BAL: So just a reminder, please speak into the mic and speak loudly. This is the loudest the system can go, so please remember to focus on that so the whole committee can hear you.

CO-CHAIR ANDERSON: Mahesh?

DR. KRISHNAN: Just a clarifying question. I know we said this yesterday. I just -- I'm trying to reconcile this in my mind. We are thinking about the evidence of the two pieces as opposed to the evidence of the totality of the measure, right? Because it's a combination as opposed to one measure.

We are looking at the evidence grading for adult Kt/V, pediatric Kt/V, separately, as opposed to the evidence that supports the use -- the combined measure.

CO-CHAIR CROOKS: This is adults only, but it's the same --

DR. KRISHNAN: Concept.

CO-CHAIR CROOKS: Same thing.

CO-CHAIR ANDERSON: Any other comments? Discussion? Are we ready for voting?

MS. BAL: Turn your mics off. Thank you.

So we are voting for evidence for 0249. The options are one, high; two, moderate; three, low; four, insufficient. And voting is open. We do need a couple more. We're at 19, and the full committee is here, so we should have 23.

Please remember to point at me. With your clicker.

Well, 20 is quorum. And if everyone has put their vote in, I don't -- okay. So the results are six high, 15 moderate, zero low, zero insufficient, and we can move forward to gap.

CO-CHAIR ANDERSON: All right.

DR. BHAN: Okay. So for the

performance gap, and I think this is where there is going to be the most attention, and we addressed this a little bit in our last discussion with the last item, so the performance data is based on 2013 CROWNWeb and Medicare claims data.

And out of -- with about 5,500 facilities, the mean performance score was 93.5 percent, with a standard deviation of seven, so crossing 100 percent actually. There were disparities data presented, which had some significance based on the large numbers, but at least I didn't feel were that dramatically impressive.

But our key comments here were -- in our call were regarding how much room for improvement there really is with a mean of 93.5 percent. I think this came up with the last HD clearance discussion as well.

CO-CHAIR CROOKS: Yes. I agree. I have concerns about the gap in the standard deviation going over 100 percent, very narrow.

Other thoughts?

DR. FISCHER: Just briefly to remind everybody the similar measure that we discussed at the end of the day drew data from USRDS. I think there was 97 percent. It was a very similar measure, just kind of furthering that up. These are slightly different data sources, probably different time periods. I don't remember which time period that was from that Paul mentioned off my head. But just going back to the point about topped out.

CO-CHAIR ANDERSON: And the QIP data shows it at 98 percent, and so it really is a measure that I think we might want to put into reserve as being a topped out measure. And, you know, there is really very little room for performance improvement.

Any further discussion? Call for the vote.

MS. BAL: Okay. So we are voting on gap for 0249. The options are one, high; two, moderate; three, low; four, insufficient. And

voting is open.

Please look at your clicker. If you're getting any sort of error, let us know.

Okay. The results are zero high, six moderate, 16 low, one insufficient. So this measure does go down on gap. We can ask the committee if they would like to consider reserve status.

Okay. I'm seeing shakes of the head as yes. So let's go ahead and do a hand vote. So if you would like to do reserve, put your hand up, please. Okay. That's a majority. So we can move forward with reliability.

DR. BHAN: Okay. So for reliability, for the specifications, the data elements were defined based on a treatment file for the patients who are on dialysis, and for testing it was performed using the data from 2013, CROWNWeb and Medicare claims with, as I mentioned, over 5,500 facilities.

The IUR was 0.942, so pretty high. Let me just see our comments. Basically, people

felt the specifications for reliability seemed pretty solid.

CO-CHAIR CROOKS: Nothing to add.

CO-CHAIR ANDERSON: Lorien?

DR. DALRYMPLE: So it was unclear to me, are four times per week hemodialysis treatments excluded? At the top it does not appear they are, but then there is denominator logic that says patient is not on frequent dialysis? So I think I would argue if you're on dialysis four times per week you shouldn't be required to need a single pool of 1.2.

DR. DAHLERUS: So that's correct. Patients on -- adult patients on four times or greater are not included. It is just thrice weekly dialysis, and we can clarify that in the specifications.

DR. DALRYMPLE: Okay.

DR. DAHLERUS: Thank you.

DR. DALRYMPLE: Thanks.

CO-CHAIR ANDERSON: Alan?

DR. KLIGER: I just want to piggyback

on that for a moment, because the specifications I think here are clear that it's three times a week. Period. Right? It is neither more nor less than three times a week, and single pool Kt/V can only be used if there is a single frequency.

We'll get to talking about that with the next measure, but for this measure my understanding is that it's a single frequency, three times a week. It could be if it's three times a week. It's frequency -- it's a frequency measure.

CO-CHAIR ANDERSON: Any further discussions or comments? Are we ready to vote on reliability?

MS. BAL: Okay. We are voting on reliability for 0249. The options are one, high; two, moderate; three, low; four, insufficient. And voting is open.

Still waiting for a couple, if everybody could try to vote again. Yes. Please vote again, just so we can get everyone.

Okay. We have four high, 18 moderate, three low -- I'm sorry, four high, 18 moderate, zero low, zero insufficient, for 0249 reliability.

We can move forward to validity.

DR. BHAN: Okay. So for validity, you know, this is based on the association of clearance with mortality. They did do some testing here looking, again, at the Spearman correlations between achieving the target and the SMR and SHR. And the coefficiencies we have seen before are -- they are statistically significant, although the magnitude is relatively small. So with SMR it was negative 0.085, and the SHR was negative 0.159.

And I think that's all the comments we had. There weren't any significant exclusions.

CO-CHAIR ANDERSON: Any discussion? All right. We'll call for the vote.

MS. BAL: Okay. Voting for validity for 0249, the options are one, high; two, moderate; three, low; four, insufficient. Voting

is open.

The results of the vote for validity for 0249 is three high, 19 moderate, one low, zero insufficient, and we'll move forward.

CO-CHAIR ANDERSON: All right. On to feasibility, Ishir?

DR. BHAN: Okay. So for feasibility, this data is in CROWNWeb, and certainly there is lots of claims data. It has been done. It seems feasible.

CO-CHAIR CROOKS: I agree.

CO-CHAIR ANDERSON: All right. Any discussion? Call for the vote.

MS. BAL: Okay. Voting for feasibility for 0249, one, high; two, moderate; three, low; four, insufficient. Voting is now open.

Okay. The results for feasibility for 0249 is 18 high, five moderate, zero low, zero insufficient, and we'll move on to usability and use.

DR. BHAN: Okay. So for usability, it

is being used. It is publicly reported. Not much more to say. I think there is not too much question here.

CO-CHAIR CROOKS: And to the gap issue, they were unable to show improvement in the last -- I think this is the last four quarters of -- I'm not sure what year, 2013, suggesting that, you know, it is consistent with the notion that it's topped out.

CO-CHAIR ANDERSON: Any further discussion on usability? All right. We'll call for the vote.

MS. BAL: Okay. The vote for usability and use of 0249 is open. The options are one, high; two, moderate; three, low; four, insufficient.

Okay. The results for usability and use for 0249 is 17 high, six moderate, zero low, zero insufficient, and we can move on to the overall vote.

DR. KRISHNAN: Can I ask a clarifying question just about the term "usability" for the

NQF staff? When we say "usability," we have often in this setting been using it as it can be used for public reporting.

Is a component of use that the dialysis unit or whoever is being evaluated can receive that data back and actually do something with it in a timely fashion? Or is it only about it can be used for a public metric? How do you balance out the operational?

So, for example, if a metric is there but the facility can't calculate it itself -- clearly not this one, but just as an example -- they can't improve because they don't know where they are. Is that part of usability? Or is that part of implementation? Or could you help me with that?

MS. BAL: I mean, I'll refer it to Sarah, but I'll let her --

MS. SAMSPEL: Yes. I mean, you know, so one of the first requirements is the fact that it is publicly reported, and that we are moving -- you know, we want the entire industry, renal,

everybody else, to be moving towards public reporting, at the same time that usability is an overall criteria even on the development phase of something that whoever is reporting on the measure can then take that measure back and use it to improve performance or improve quality in some way.

So, yes, that is one of the things that you should be considering under use and usability.

Did you have anything you wanted to add, Marcia?

DR. KRISHNAN: So just so I can clarify, it's the ability to get real-time access to the data, so whoever you're evaluating can say, "Oh, my gosh, I'm not performing well. I've got to improve."

MS. SAMSPEL: Well, it's not necessarily real time.

DR. KRISHNAN: Not real time. I keep using the wrong word, and we keep clarifying.

MS. SAMSPEL: The ideal situation is

you would have real time, but, you know --

DR. KRISHNAN: The facts on the data.

MS. SAMSPEL: -- even if you're talking about HEDIS, you know, it is something that should be coming back that has an opportunity for the organization to use to improve quality.

DR. KRISHNAN: Okay. Yes. I apologize. I keep using the wrong words, but we have different terminologies.

CO-CHAIR ANDERSON: Any further discussion before we call for the vote for recommendation for endorsement for reserve status? All right.

MS. BAL: Okay. And as a reminder, the specifications for reserve status is that it is topped out, but it would be harmful to lose the measure, and that it is very strong in reliability and validity.

So the vote for reserve -- for endorsement meets potential for reserve status for 0249. The options are one, yes; two, no.

Okay. The results are 22 yes, zero no, so this has a potential for reserve status, and we can move on to the next measure.

CO-CHAIR ANDERSON: Okay. All right. Measure 1423, and, well, you guys are ensconced there, Claudia and Joel?

DR. DAHLERUS: So this is a measure of pediatric hemodialysis adequacy based on Kt/V. It's a percentage of patient-months that achieve Kt/V 1.2 or more. And this was developed in 2010 by the pediatric hemodialysis adequacy TEP, and it was originally endorsed in 2011.

The underlying context of the TEP's recommendation recognizing that there was not a lot of -- a large body of evidence defining a specific threshold for the pediatric population, but they did cite the KDOQI, the 2006 KDOQI guidelines, which recommended -- and these were opinion-based, expert opinion, but the guidelines recommended that for pediatric patients Kt/V should align with the minimum dialysis adequacy for adults of 1.2. And so that was the rationale

for the TEP defining this threshold.

Thank you.

CO-CHAIR ANDERSON: All right. And it's Alan and Frank?

DR. KLIGER: So as we just heard, it is a measure for pediatric hemodialysis, and the specifications are for patients dialyzing three or four times a week whose average delivered dose of hemodialysis using UKM or Daugirdas II are a single pool Kt/V of 1.2.

So I want to ask first a clarifying question of the developers. The three or four times a week, is that what was in the original 2011 version that was approved?

DR. DAHLERUS: Yes, it was. And that was also stated in the TEP's formal recommendations. The rationale, as I understand them, is they wanted to ensure that children that are dialyzing four times a week would still be evaluated for adequate dialysis.

DR. KLIGER: Yes. So we will go through the evidence in each of the categories,

but I want to lay out from the outset what I think is a fundamental flaw and problem in this measure. And it is -- it will apply for this measure and others as well.

The first, which we will deal with quickly, is that as in many pediatric measures, we don't have very much direct evidence for kids. And so this is really based on adult data with the assumption that children should be doing at least as well as adults do, and that's, I believe, a very reasonable position to take.

The problem is a kinetic measurement problem, the way that the measure is currently constructed. And just because I wanted to make sure that I had this right, I spoke to John Daugirdas, the kineticist who constructed the Daugirdas II formula to make sure that what I was saying and what I am going to be sharing today is accurate.

When either the urea kinetic measurements overall or specifically Daugirdas worked with his group in 2003, and then

eventually made the recommendations to KDOQI in 2006, the challenge was to come up with a measure that somehow approximated what the kidney really does. Kidney works 24 hours a day, seven days a week, obviously.

But in hemodialysis that is intermittent treatment. The dynamics are very different, because you have a burst of clearance and then a long period between them of no clearance at all.

And so a formula that measures what happens in one specific treatment will be very dependent on what happens around it, how frequent the treatments were. So, for example, in the measure as it is currently constructed, if a child is dialyzed three times a week, and the single pool Kt/V is measured -- and, of course, the way it's done is a single treatment is assessed with pre- and post-dialysis urea, and you calculate the Kt/V. You'll have a number, whatever it is. If that same child is increased to four times a week, or five times a week, using

the methodology that is here, the Kt/V will not change. It will be exactly the same because of the way it is calculated.

And the kineticists have been clear, and they were back in 2006, that if we are looking at varying frequencies of dialysis that rather than using a single pool Kt/V, the tool that we should be using is a continuous tool, one like what they have called the standard Kt/V.

And, in fact, in the KDOQI document back in 2006, they have a table, Table 18, in the Appendix that Dr. Daugirdas referred me to that shows what the relative clearance would be required depending on frequency.

So, for example, the minimum Kt/V of 1.2, giving patients who are dialyzed an average of 3.5 hours a week, is where we currently sit for three times a week dialysis. If you look at patients who are dialyzed four times a week, that same point would not be 1.2. It would be 0.77.

So if we were looking at adults or kids, all dialyzed four times a week, the minimum

dose that we would be asking for is not 1.2. It would be .77, but that's only if all the patients were dialyzed four times a week.

So my point is that using single pool Kt/V in patients dialyzed at different frequencies is like using a tape measure to measure somebody's weight. It is the wrong tool, and it really cannot consistently be used. And Dr. Daugirdas was clear with me that he felt passionately about that as well.

The rejoinder could be that, well, that is okay, as long as they're achieving that minimum of 1.2, if they're dialyzed four or five times a week, then they're way over the threshold anyway. The problem with that, I believe, is that by setting up a minimum of 1.2 with whatever frequency it would be a disincentive to put patients on increasing frequency of dialysis, because that wouldn't change their Kt/V.

So a large child with a Kt/V of 1.0, done three times a week, with this measure, if you went to four times a week it wouldn't change,

or five times a week it wouldn't change. So the incentive would be instead of going to more frequent dialysis to extend the time, which -- during single sessions, which if you do the calculation is not nearly as efficient a treatment as increasing the frequency would be.

So my point I guess is that there is, I believe, a fundamental flaw when we use this tool, if we are looking at varying frequencies of dialysis. And in the real world we now know that varying frequencies are becoming more and more common, not only for kids but for adults as well. So I believe that that is really a fundamental flaw in this particular measure.

The second thing that I will mention, but we have -- I mentioned yesterday as well is that for hemodialysis we are not including endogenous kidney function. And it is reasonable to do that, I believe, since most patients in any case lose virtually all of their endogenous kidney function within the first year.

But I would love to see these measures

allow for -- not require, but allow for including endogenous function in this measurement. So that incident patients with substantial kidney function still available could have crafted therapy to give them less dialysis time initially until their endogenous kidney function was reduced to the point that they required more treatment. So that is just my general comment.

When we go now specifically to the evidence, the evidence here is based on the adult experience. We have heard already the evidence for the adult experience, which is strong for three times a week, Kt/V of 1.2. And the statement in KDOQI, as we have heard, is that kids should have at least as good treatment as adults. So that is where the evidence is.

Now, I don't know, Connie, whether you think we should vote on sort of the flaw in the design, at least my perception of the flaw of the design, in the evidence phase or in the specifications phase.

CO-CHAIR ANDERSON: Well, I think we

would be talking about it in the specification phase.

DR. KLIGER: Okay.

CO-CHAIR ANDERSON: I would like to hear from our pediatric experts over here.

MS. BAL: If you wouldn't mind, could we also discuss it during the specification? Just so we can make sure that, you know, what people are hearing is focused on evidence, and people vote on that, and then we'll move to the specifications.

CO-CHAIR CROOKS: And it's a tough overlap between where it should be I guess. I have trouble voting on the evidence if this is an issue. So I'd prefer to get it talked about now.

DR. SOMERS: In terms of evidence, although most of the evidence was within the KDOQI, there was some CPM data that pertained to adolescents particularly that showed that Kt/V less than 1.2 is associated with a greater risk of hospitalization. There is some other literature from Europe that supports that as

well.

CO-CHAIR ANDERSON: Lorien?

DR. DALRYMPLE: But with respect to our evidence voting, to my knowledge most of the evidence -- but this is where the pediatric nephrologist would be very helpful -- relates to three times a week assumptions. So if it is specified on three or four times a week, doesn't it fail on evidence on the four times a week frequency? Because the evidence doesn't support that frequency unless our pediatric nephrologist can inform us.

CO-CHAIR ANDERSON: Josh?

DR. ZARITSKY: I mean, you would have a hard time, even if you just substitute "pediatric" for "adult," and you said, well, you're going to have four times a week or daily dialysis, and use a cutoff, a standard single pool Kt/V of 1.2.

So I don't -- I have a problem with using it for four times a week. If they are going to use a standardized Kt/V or other than

single pool, I think that would be different, and we get into the evidence of a standardized Kt/V, but -- and there are quite a few children that are dialyzed more than three times a week for fluid issues as well.

I will tell you, practicality, it is usually fairly easy to achieve a Kt/V above the standard in pediatrics, but the four times -- just on a technical basis, I have the same type of feeling about the inclusion of the four times a week.

CO-CHAIR ANDERSON: Lori?

DR. SOMERS: But what Alan said would still pertain. If you are meeting a Kt/V of 1.2 four times a week, then you obviously would be avoiding all the bad sequelae of having poor adequacy with your dialysis. And my understanding of kind of the proportion of children who are dialyzed four times a week is somewhere between five and six percent, so it is, you know, a relatively small number of kids at this point in time.

MS. HARTWELL: I just wanted to piggyback on Alan's comment. It kind of frustrates me to hear that, you know, the residual kidney function is a secondary issue, and to the patient that is the primary goal, I mean, of any patient on dialysis is to keep their residual kidney function. It makes such a difference in their quality of life, their diet, and to think that a measure doesn't help that or maintain that is extremely frustrating to hear.

And I think that that should be in consideration when a developer creates a measure, because we don't -- first, we don't want to do any harm. And if a physician can help keep the kidney function, and, for example, when I was on Next Stage, I had quite a bit of kidney function, and so I only needed it three times a week. And that went on for about six to eight months because I was able to maintain my own kidney function.

But what I'm hearing is is that may not always be the case, if a doctor is

prescribing -- you know, so that's just my comments. So I'm a little concerned about hearing that.

DR. KASKEL: So in the small infant and small child, it is very common to be doing dialysis at least four times a week. And even if you achieve the goal and they are not growing, yet your nutritional assessment is adequate, and most of these infants are on G tubes, so they are getting the nutrition at night, we increase the dialysis time. We'll go to five days a week. We have infants on dialysis sometimes five days a week, most of them four, and we measure growth. If they don't grow, they may need more, despite the calculations.

So it's a very soft sign, and it takes a long time to determine growth. It's not just a week; you're looking at months.

CO-CHAIR ANDERSON: Frank?

DR. MADDUX: So as we talked about this in the workgroup, and we bought up yesterday, one of the issues with a number of

these, and certainly with this one, with the small body mass that patients have, and the others, is whether the urea kinetic model, or simply the Daugirdas formula, really equate to the same thing.

I would ask my pediatric colleagues whether the urea kinetic model that was predominantly built in adults is 100 percent applicable, especially when you get down to the very smaller sized children, especially looking at the urea generation rates and the nutritional issues that I think were just brought up.

And these are just issues where we are applying something to a population that really was developed for a different population. I think we need to make sure we are aware of that as we look at this one.

DR. KASKEL: I wouldn't go too far into the urea modeling or kinetics for us in the growing -- you know, maybe a stable adolescent who is not sick and losing weight. That's one thing. But not in the growing child or the not

growing child or adequately growing child, no, I wouldn't.

DR. MADDUX: One other corollary question on that for you all is, do you even do formal urea kinetic modeling often in children?

DR. KASKEL: I think that we are -- you know, we are stuck, once again, with the tools that -- there is a limited set of tools, and so we do use this as a tool. So I do think it's still an important measure for us to be calculating it and to -- especially as, you know, our children grow.

But, obviously, we take a lot of other things into consideration, but we just don't -- you know, we don't have the body of evidence. And, you know, you could make the argument it's small solute -- you know, is there just overemphasis on small solute calculations because they're easy to do? We know the modeling, but they're still a tool that you have to use, and maybe that's the best tool we have now. So it's still a very valuable tool for us.

CO-CHAIR ANDERSON: Alan?

DR. KLIGER: So let me ask you what tool you are using. Are you using the standard Kt/V or the single pool Kt/V when you have patients that have varying frequencies of dialysis?

DR. SOMERS: Standard for varying.

DR. KLIGER: You're using the standard Kt/V, not the single pool.

DR. SOMERS: At least that's what I'm using.

CO-CHAIR ANDERSON: Any further discussions or comments? All right. We are ready for the vote on evidence.

MS. OGUNGBEMI: The committee is now voting on evidence for Measure 1423. The options are one, high; two, moderate; three, low; and, four, insufficient. Voting is open.

The results are zero votes high, nine votes moderate, 11 votes low, and three votes insufficient. The measure does not fail on evidence. This is Measure 1423. It does fail,

I'm sorry, does not pass on evidence.

MS. SAMSPEL: Actually, and so what we want to do here, and something that we have afforded to other developers, is -- and what a fail means at this point is that we do not move forward with the vote.

However, what we would ask, and I think Alan has already expressed, you know, some of his concerns, but also any additional advice that you can provide to the developers as they consider this measure, if they do take it back to a TEP, et cetera, what are those types of things that would change the game on this measure and how you would vote in the future.

DR. ZARITSKY: Just a procedural -- are we allowed to move forward with the other consideration, that there is not enough evidence but we are still -- what is that fourth -- the exception? Or that has to be -- then they have to have a majority of insufficient evidence vote. Can you remind me of the procedure there?

MS. BAL: Yes. We can still move

forward and vote on insufficient with exception, if that's what the committee would like to do.

DR. KLIGER: Can I just raise an understanding of that? Because if it's approved with exception, doesn't that say that we think that it's a valid measure, it's just that it doesn't -- it either has topped out or there is some other reason not to be using it, but that has strong evidence for it? That's the way you have instructed us, that there is strong evidence for it.

MS. BAL: No. For the exception with -- the evidence with exception -- I'm sorry, insufficient evidence with exception, it is that there is low or insufficient evidence to support the measure; however, you feel that other factors are strong enough to move this measure forward, such as you think that this is an important enough measure, or that the other factors, such as reliability, or the fact that it is in use and it is being used well, that kind of thing.

So if you think that other factors

will be higher, and this is a measure that should be kept, then yes. The exception is on the evidence, so if evidence would be low, everything else you should consider would be higher.

DR. KLIGER: I would just argue I guess that it means that that would be an in-practice measure. And I think this is really a fatally flawed measure as currently constructed. I think if the team goes back and substitutes standard Kt/V, and does the work to do the measure -- you know, the background work to do the gap analysis, et cetera, using standard, then we might feel differently. But I would -- I just have trouble approving this with exception in its current form.

DR. SOMERS: I mean, or we could just restrict it to children being dialyzed three times a week because it's such a small proportion of kids who have more than three times a week.

CO-CHAIR CROOKS: Well, we can't do that. The developers could take that --

DR. SOMERS: Right. No, but that's

our suggestion.

CO-CHAIR CROOKS: Yes. I'm thinking of the exception as, for instance, you know, we know it's important; the evidence just isn't there, like jumping out of a plane without a parachute is a bad thing, but no one has ever done the trial. I think we would give that an exception. That's the sort of -- but here we are saying that the -- we are not sure it's that important, and as it's constructed it has flaws.

But I think -- anybody else who would like to speak to considering this as an exception?

DR. ZARITSKY: So with just that in mind, I mean, you know, I think that this is, you know, an important measure, and I would like to see the four times a week -- I would rather not see a standardized Kt/V. I would rather see the three times a week, you know, single pool implemented.

And I brought up the exception issue because I do think it's an important -- it's an

important measure that we have and that we don't just totally, you know, get rid of it. It's --

CO-CHAIR ANDERSON: We can call for the vote for exception.

DR. ZARITSKY: So just in the future, if you're going to -- if you have an idea that you want to do an exception, then the category you should be voting for is, then, number four, right? Yes?

CO-CHAIR ANDERSON: Any more discussion in terms of calling for the vote for exception?

DR. FISCHER: Can you just clarify what we are going to vote on? The exception of -- maybe I'm just a little bit confused. The exception of what? What exactly are we voting on, please?

MS. SAMSPEL: So basically, you know, when the votes fall down into this three or four category, so low or insufficient evidence, typically a high number of votes in insufficient would mean that you think that while the evidence

provided did not support the measure that it is still an important enough measure that you would like to see it continue through the voting and recommend endorsement with exception to the evidence.

You know, I think in this case, you know, it also sounds like there was some confusion, how do you get to that vote of exception for evidence, is it low or is it insufficient, and so, you know, I think what I have heard is that there may be a consideration, and we might just want to get out on the table, do the vote of how many people want to say, you know, we think this is an important enough measure to continue to move it forward through the voting process.

I think the other things that I keep in mind here are the fact that this is all going to be documented in the report, that there is also this opportunity during public comment to get more information out of the public on how they feel about the measure that will come back

for your consideration as well.

So the measure doesn't technically die here in this room today if it does not go all the way through.

DR. KRISHNAN: But just to what Josh said, Sarah, if he wants to have the exception -- or if he wants to have the measure only apply for three times a week, am I voting on what is written, or do we have to go back and --

MS. SAMSPEL: In this case, we are voting on what is written based on --

DR. KRISHNAN: Okay. So without the three times a week.

MS. SAMSPEL: Correct.

DR. KRISHNAN: Okay.

DR. FISCHER: So if we vote yes, then regardless of concerns about evidence we are going to move forward and discuss the measure and potentially move to endorse it. If you vote no, then you feel that it's not -- regardless of the evidence, you don't feel it's worth further discussion. Is that correct?

MS. SAMSPEL: So I -- I mean, if you vote yes with exception, we will continue with both the discussion and the vote, you know, as long as it then passes reliability and validity. In the event that you vote no, that you don't feel that it needs exception, then, you know, technically the votes stop, you could still discuss and give some recommendation, so that we have it in the report, and that goes out to public comment.

We work with the developers in the meantime to see if they can make revisions prior to the post-comment call.

DR. KASKEL: I think it's important enough with all of the flaws that Alan so beautifully brought up, is that this should be an exception, and we don't -- we used this, and we need this. It's a guideline. I don't want it to sound like we don't use it. We do.

Small percentage of patients may need dialysis more than three times a week because of growth issues and, et cetera, but I still think

it's all we've got. And, yes, there needs to be some development, some TEP, and input, to make it better.

CO-CHAIR ANDERSON: Andy?

DR. NARVA: No. That was exactly my question. I -- because if we -- this goes away, we have no pediatric adequacy measure. And I realize we are only supposed to talk about evidence, but my sense is that pediatric nephrologists want this as a tool to improve care, despite its flaws, and that would be correct.

CO-CHAIR CROOKS: There is a later measure coming up that does include this as kind of a compilation of the four. And does that have the same specification? It does.

DR. ANDRESS: There are two more measures. One is a combined pediatric and adult human dialysis measure, and it is a combination of the adult and human dialysis measure that was just discussed and this measure.

And then there is the composite

measure, which -- well, the overall measure that includes the human dialysis and the peritoneal dialysis measure that we discussed yesterday. And that would include the same specifications for pediatric patients on human dialysis.

CO-CHAIR CROOKS: So I think we need to discuss and vote on whether we want to consider this an exception, this other --

DR. SOMERS: The other measures, the frequency for children can still be four times a week. Is that correct? I thought that was the case.

DR. ANDRESS: As currently specified, that is correct.

DR. SOMERS: But the same issue.

DR. KLIGER: That would be the same issue for children and adults.

DR. DAHLERUS: So the three or four times was only specified for the pediatric measure. It is not -- has not been specified for the adult, and, again, that change would be reflected in the other measures.

DR. KRISHNAN: Can we do what we did yesterday for the upper bound gate thing? Can we do -- it's the exception but with the stipulation of this exclusion?

CO-CHAIR CROOKS: I don't view it parallel because it -- in that case, it just -- it didn't change a thing except to remove it. But here the evidence base is changed. We don't know should they be going to standard or to single pool only three times a week, and there's other things that the developer would need to consider, I believe.

CO-CHAIR ANDERSON: Frank?

DR. MADDUX: So I just want to be clear, again, to Michael's point, what we're going to be voting on here. It strikes me that -- I'm a little concerned about the degree of the way we view stipulations. I certainly understand it, and think for the typos and the administrative sort of logistics item that is fine. But to me this is much more fundamental.

And my assumption is, recognizing that

this is an existing measure that is in use, yet a measure that has some distinct flaws because of the three or four times, is the exception for us to say that this measure, as designed, would still be okay, versus this measure needs reconfiguring to be designed properly in either a future iteration or an adjustment?

CO-CHAIR CROOKS: I would take it to mean that voting for an exception means that we think the measure is okay as it is, and not necessarily recommending a reconfiguration, whereas if we don't give an exception, they do have the opportunity to -- that almost pushes them to make it better somehow.

CO-CHAIR ANDERSON: Josh?

DR. ZARITSKY: Just for the sake of the exception, which I would recommend that we at least put it up to a little vote, is that I could make the argument that the evidence, as it stands, even for a 1.2 Kt/V, there is very limited evidence, where, again, borrowing from the adults, and so I think that's another avenue.

So to get hung up on the -- you know, the four times a week when it's five percent, and we really don't know, is that really 1.2, maybe if we do 1.2 for four times a week there is additional benefit.

You know, technically I have a problem with it, but philosophically I would like to see it put up for the exception.

CO-CHAIR ANDERSON: Alexandra?

CO-CHAIR CROOKS: Lisa is first.

DR. LATTS: So I am struggling with hearing the evidence here versus hearing our pediatric colleagues who know more -- way more about this than I will ever ever know, telling me they need this measure.

And then I am also struggling with there are other measures that we will need to reconcile at some point. So I guess I'd like to see this one go through the process, then reconcile and we sort it out.

DR. DALRYMPLE: Can I just ask a procedural question? Because I think it's clear

we all want a pediatric measure, but we want it to be valid with kid specifications.

(Laughter.)

So just for our own understanding, Sarah, suppose we decide not to pass making exception for insufficient evidence. So we decide today that the majority does not go that way, in theory, now that we are a standing committee, how quickly could CMS bring a new measure back to us, us reconvene via telephone workgroups, if they followed, for example, the pediatric recommendations saying just -- just make it three times a week for it to have evidence that supports -- I mean, how quickly could it come back to us for review and make it to public comment, et cetera?

MS. SAMSPEL: So, I mean, procedurally what would happen in this case for this measure is that if you fail it here, you know, again, technically the votes stop, you could still continue your discussion for sake of the future measures as well, and that we have it in line and

also have direct feedback to the developers on what would need to be done in the future for you all to want to reconsider it.

However, if it -- basically, a failed vote here means you do not recommend it for endorsement. It still goes out for public comment, so -- and then -- so that, you know, it takes us a month to get the report together, another month for public comment, we get you all back together. During that time, if CMS can make those adjustments, you know, and provide the data, et cetera, that you would need to put it through, you could revote on it technically in a couple months.

However, you know, that is kind of an ideal situation. If that doesn't happen, then, you know, I think it's up to CMS to say these are significant -- you know, these are significant changes to the measure, and, you know, it is between CMS and NQF to determine, is it a lost endorsement or is it something that holds for a while while they have a plan in place to resubmit

the measure with significant changes.

CO-CHAIR ANDERSON: Yes?

DR. ANDRESS: So in terms of timeline and responding, I agree, I think this is a somewhat different circumstance than the measures discussed yesterday, in part because I think the stipulation there was a fairly straightforward case of not changing the measures, and we knew what the answer was at the time, without changing the measure.

In this case, of course, there is the potential that the measure itself is changing. You know, is the answer -- you know, just a question I would have for -- to look at it, is the answer to retain patients who are treated four times weekly with a different target, or is it a different way of -- should there be a different way of calculating target, or just simply to remove the four times weekly patients entirely? I mean, I think that's a question to be investigated and not something to be stipulated to in this meeting.

In terms of the timeline, I think we are pretty confident that we can investigate that and get back to the committee following the public comment period, you know, and that can be considered.

I think the other issue to consider is that there is -- we are planning on a renal project next year as well, where we are already going to be submitting some other measures, and keeping in mind of course that this -- this probably rebounds to the two following measures as well. That might also be -- if we can't resolve it in the two months until the public comment period is over, that might be another opportunity to address the concerns.

I think from our perspective of course we want to retain pediatric patients in the assessment of adequacy for a number of reasons that are not entirely related to evidence. I think from a policy perspective it is important to keep a focus on the treatment of pediatric patients, and they are already systematically

excluded from a number of other quality measures in our various programs.

And I think that is certainly something that is worth pursuing as this committee sees fit, and we are certainly willing to put the effort into ensuring that the measure both comports with the quality standards but also addresses what I think is a continuing need for quality in the dialysis community.

MS. SAMSPEL: Did that answer your question? I mean, do you -- does everybody feel clear on what the timelines could be based on the vote? Yes. Okay.

CO-CHAIR ANDERSON: Are we ready to make a vote on insufficient evidence with exception or no exception?

MS. OGUNGBEMI: The committee is now voting on Measure 1423, evidence, potential exception to empirical evidence. The options are one, insufficient evidence with exception; and, two, no exception. The vote is now open, and I will be voting for Ms. Hartwell, since she

stepped out of the room.

The results are 14 votes insufficient evidence with exception, and nine votes no exception. The measure passes with insufficient evidence with exception.

CO-CHAIR ANDERSON: All right. So let's move on to performance gap. Alan and Frank.

DR. KLIGER: Okay. Performance gap, the developers did look at a relatively small group of patients. Congratulations, you found the pediatric patients to examine for this in the CROWNWeb 2013 data set showing 85 percent adherence with a range between 56 and 100 percent. There were no -- not enough evidence to look at skewedness or subgroup analysis.

DR. MADDUX: The only thing I would add is that there was a mild gap recognized in the small population that was studied.

CO-CHAIR ANDERSON: Any further discussion? Are we ready to vote on performance

gap?

MS. OGUNGBEMI: The committee is now voting on performance gap for Measure 1423. The options are one, high; two, moderate; three, low; and, four, insufficient. Voting is open.

The results are one vote high, 17 votes moderate, two votes low, and two votes insufficient. Measure 1423 passes on performance gap.

CO-CHAIR ANDERSON: Moving on to reliability.

DR. KLIGER: Reliability of the specifications, I guess we've already beat that dead horse. So I think that there is a real problem with the specifications and personally would find trouble passing this on the specifications.

In terms of the reliability testing, there was, again, testing on this small group of patients showing an IUR of 0.81, saying that about 81 percent of the variation of the measure is attributed to between facility variation. The

pre-reviewers, who incidentally did a fabulous job -- thank you very much -- found that this translated to moderate reliability, and I agree with that.

CO-CHAIR ANDERSON: Any further discussion on reliability? All right. Ready to call for the vote?

MS. OGUNGBEMI: The committee is now voting for Measure 1423, reliability. The options are one, high; two, moderate; three, low; and, four, insufficient. Voting is open.

The results are one vote high, 11 votes moderate, seven votes low, and two votes insufficient. Measure 1423 fails on reliability. Possibly not. One moment. Yes, it -- is it --

MS. BAL: So we would need a high moderate to be at least 60, so it's actually gray zone. Sorry, I shouldn't have said failed. It is a gray zone measure. So for -- I think we briefly went over it through one of the workgroup calls, but for anyone who was not on the call, the gray zone is a place between 40 and 60

percent. So that's the point we're at. We're at 57 percent of agreement.

So basically what that means is that we were not able to come to consensus on reliability, and we would open that up -- when we do our draft report, we will write that in the draft report and say, "Ask for comments from the public," but we do move forward with the measure. But we basically said the committee didn't come to consensus.

DR. FISCHER: People didn't vote. Two people --

MS. BAL: Lori is not in the room. So, and we can't vote for her. Well, she wasn't at her -- we can do a revote and --

DR. FISCHER: Yes. I'm just pointing it out because we didn't have --

(Simultaneous speaking.)

MS. BAL: And please check your batteries, if we're not getting your vote.

DR. FISCHER: So we're voting on -- just so I'm clear, we're voting on both the

specifications and the reliability. So it's not just about the reliability of testing. It's also about the specification, which have been -- I'm just confirming, just so I understand what we're voting on again.

CO-CHAIR CROOKS: Yes, you're correct. We're voting on both, which may or may not have -- be the right way to do it, but --

MS. BAL: So we will go ahead and revote, and just, again, to clarify, to answer Michael's questions, when you're voting on reliability, you are voting on both 2(a)(1), which is the precise specifications, and the testing, so the appropriate method and scope with adequate results of the testing.

The other thing I am just going to ask is that when you vote, and you point at Alexandra, that you make sure your number is showing up in your little screen. And I know people are, but this is really kind of a really important place where we need all the votes.

MS. OGUNGBEMI: The committee is now

voting for Measure 1423 on reliability. Options are one, high; two, moderate; three, low; four, insufficient. Voting is open.

MS. BAL: As you've seen, we do need those two other votes.

MS. OGUNGBEMI: The results are zero high, 11 votes moderate, 11 votes low, and one vote insufficient. The measure is still in the gray zone. So we will continue with the next criteria.

CO-CHAIR ANDERSON: All right. Moving on to validity.

DR. KLIGER: This one is easy. It's on face validity, so that is the criteria.

CO-CHAIR ANDERSON: Frank, any other comments?

DR. MADDUX: No other comments.

CO-CHAIR ANDERSON: Any further discussion? Are we ready to vote for validity?

MS. OGUNGBEMI: The committee is now voting on validity for Measure 1423. The options are one, high; two, moderate; three, low; and,

four, insufficient. Voting is open.

Results are zero votes high, 18 votes moderate, four votes low, and one vote insufficient. Measure 1423 passes on validity.

CO-CHAIR ANDERSON: Moving on to feasibility?

DR. KLIGER: It's used in CROWNWeb. It has been used. I don't whether we can do both of them together, but it has been used and it appears to be feasible.

CO-CHAIR ANDERSON: Any further comments or discussion? Call for the vote.

MS. OGUNGBEMI: The committee is now voting on feasibility for Measure 1423. Options are one, high; two, moderate; three, low; and, four, insufficient. Voting is open.

Results are 14 votes high, eight votes moderate, one vote low, and zero votes insufficient. Measure 1423 passes on feasibility.

CO-CHAIR ANDERSON: Usability and -- use and usability?

DR. KLIGER: So it's currently reported in Dialysis Facility Compare, and the QIP, so it appears to be usable.

CO-CHAIR ANDERSON: Any other further comments or discussion? Call for the vote?

MS. OGUNGBEMI: The committee is now voting for Measure 1423 on usability and use. Options are one, high; two, moderate; three, low; and, four, insufficient. Voting is open.

Results are 17 votes high, six votes moderate, zero votes low, and zero votes insufficient. Measure 1423 passes on usability and use.

CO-CHAIR ANDERSON: Before we move on to endorsement, is there any further discussion regarding this measure, or comments or questions for the developers?

All right. We are ready to move on to vote for endorsement, and it is with exception. Is that how we state it?

MS. BAL: The exception is only for evidence.

CO-CHAIR ANDERSON: Oh, just for evidence. Okay. To move on recommendation for endorsement.

MS. SAMSPEL: Actually, overall suitability for endorsement. Sorry. And I'm just making a point about this. So you are recommending endorsement. Technically, the endorsement body is the board of NQF. So yours is a recommendation for suitability for endorsement.

DR. LATTS: Okay. So, wait, so with all of the previous discussion and the exception/approval of the evidence, when we are doing the endorsement, is it forgetting all of that and just taking it as it is written? Or is it understanding that there is this exception and we have asked the developers to go back and rewrite it?

MS. SAMSPEL: So, in this case, there is the exception, and then there is also the gray zone. So this measure has work to do, and you will have an option of revoting.

MS. BAL: Just a clarification. The exception is not that you are telling the developer to do so. The exception is just stating that you take -- you have seen the evidence, you don't feel that it's strong enough to support it, but you feel the measure is important enough, and you are going to -- where normally you would have to stop at evidence, you are saying that you fully can bypass that.

So your -- the exception was not that the developer has to make this change. However, you have given that recommendation to the developer, and they can take that and provide changes for you. I just want to make that clear, that there was an exception.

CO-CHAIR CROOKS: And what is the ultimate impact of the gray zone vote?

MS. BAL: So the gray zone, since it was only on reliability at this point, we would send it -- you know, put it in the draft report, ask for public comment. You can choose to revote, if you want. If you feel that whatever

you receive is not enough for you to revote, we would leave it as is. The only vote you would need to reconsider is if on this vote, overall suitability for endorsement, you were in the gray zone.

So same thing, we would, you know, put in the draft report, ask for comments, and then we would ask you to revote at that point to get -- see if we can get you to come to consensus, only if this vote goes that way.

DR. ZARITSKY: So it seems that it is a safe -- I mean, it would be sort of a compromise, then, if we had the option of saying we want to revote on that gray zone depending on they had -- they had made a change, right? That would give us -- that would say we have to revote on that to then pass it on the measures. Is that my understanding?

MS. BAL: Well, you can revote either way. So you can have the gray zone and revote, or you can just say, "Listen, based on what we have received from the developer, and what we

have received from public comment, we feel that we need to revote." And then you can make that decision as well. So does it -- you have that opportunity no matter what.

DR. ZARITSKY: Are we voting a revote -- I mean, I guess what I'm getting at is, how do we know that we are going to -- I think people in the room would feel a lot more comfortable if there was a revote on that gray zone as a safety measure to say, "Hey, you know, I'm not comfortable with the way that it's written now." I guess that's the question I'm getting at.

MS. BAL: We generally just do a verbal disclosure, and then we'll vote on that. And, unfortunately, there is no way for us to trigger a gray zone, but we can definitely make this a guaranteed item on our agenda for the post-meeting call. So we would discuss this and you would get to decide if you want to revote on this measure or not.

CO-CHAIR CROOKS: I still am not

exactly clear on process. In order for us to recommend for endorsement, we have to have 60 percent or greater, especially on the top two, importance and scientific acceptability of the measure. And if we haven't achieved that, then we haven't reached consensus, and how can we be voting to endorse it? Does that make -- so I'm asking if the process is that we must reach consensus on each category in order to be voting for endorsement. And we haven't reached consensus on reliability, so how can we be voting for it? So the vote is we're going to endorse it if the gray zone can be cleared, or how can we even --

MS. SAMSPEL: So this is in one way why I was trying to make that distinguishment. So in this case, basically what you're saying is overall for Measure 1423 -- that's where we are, correct? So for Measure 1423, you have actually passed it through on importance because -- with an exception to the evidence. You're in the gray zone, so you did not reach consensus on

reliability.

However, it did meet -- and you did pass it through on validity, feasibility, usability and use. This overall suitability for endorsement is then looking at a whole in all of the criteria. Would you currently have your recommendation be this measure is suitable for endorsement as it is currently specified in front of you?

Either way, what will happen with the gray zone is during -- you know, we will be seeking public comment. The developers have also a period of time to make changes to the measures. That all comes back to you during the public comment call, you know.

Kind of to address your question, Josh, it is actually up to the committee. If we got no comments, if the developer said, "Forget it. I'm not going to do anything. We're not going to put investment into this." You could say, "No, we're not going to revote. We want to leave our vote as stands." Or you could say,

"No. You know, we want to revote," and then you turn it down.

So technically this is an initial indication that right now you think this measure is recommendable to continue its endorsement process.

DR. LATTS: And that is if we vote yes. If we vote yes, you're saying --

MS. SAMSPEL: If you vote yes. Or you could vote no, and still -- there is still that almost a cure period of the public comment where you get more -- a broader range of information from NQF members and the public as well as what the developer may or may not plan to do, and you would take that into consideration. And on the call, we would say, "Okay. With all this new information, do you want to vote or revote?" You know, I'm guessing you guys would say yes, and then we would do a revote and it would be an online revote --

DR. LATTS: So it's almost --

MS. SAMSPEL: -- of all criteria.

DR. LATTS: So this vote is essentially not binding right now, one way or the other.

MS. SAMSPEL: Well, it's -- well, yes, I mean, I think it's binding in the fact that it puts it through the NQF process. It helps us understand where it is slated, what additional information we need to gather, how we put it into the report, what your issues were, bringing to the attention some of the important concepts that you brought up.

You know, the reality is is even after that final public comment vote, you know, the measures go through the CSAC, and the CSAC is an additional layer before it goes to the board.

DR. ANDRESS: I'm sorry. One clarification I would ask for. My prior understanding had been that if we -- if the 40 percent threshold is not met for the gray zone, then the measure does not continue through public comments. Is that correct? Or will it go through public comments regardless?

MS. SAMSPEL: Everything goes through public comment.

DR. ANDRESS: Okay. All right.

MS. SAMSPEL: It's just the nomenclature on the measure.

DR. ANDRESS: Okay.

MS. BAL: You may be confusing that with voting. We don't vote on measures that don't go through, but we would still do public comment. I'm sorry, not -- the committee would vote. The membership does not vote on those measures. I think that's new here.

CO-CHAIR ANDERSON: All right. Are we ready to vote for suitability for endorsement?

MS. OGUNGBEMI: The committee is now voting for overall suitability for endorsement on Measure 1423. The options are one, yes; two, no. Please vote.

MS. BAL: So the measure is in gray zone, so you will get to vote on it at the post-meeting comment -- I mean, the post-comment meeting.

CO-CHAIR CROOKS: How are you enjoying the land of Kt/V today? It just never ends.

Okay. So we are moving on to Measure 2703. I'm going to take the helm as Connie is recused on this particular --

CO-CHAIR ANDERSON: No. I'm not recused. I'm evaluating.

CO-CHAIR CROOKS: There's a conflict? Oh, it's not this one.

CO-CHAIR ANDERSON: No.

CO-CHAIR CROOKS: Oh. Okay. Well, then, you may continue. You're doing a great job. I guess I'll -- okay. This is Measure -- CMS measure minimum delivered dose, hemodialysis dose. Our developers are at the table, and take it away.

DR. DAHLERUS: And, again, I will keep our remarks very brief. So this is the combined hemodialysis adequacy measure. So it includes both the adult and pediatric populations, and the assessment is whether the respective adult and pediatric populations achieved minimum Kt/V, and

that in this case is 1.2.

As we stated earlier, and in yesterday's discussion, we will be removing the upper threshold of 5.0.

DR. MADDUX: So we won't go through the entire discussion we just went through, but basically this is the combined measure. For the evidence, the KDOQI guidelines were used from 2006. The guidelines related to pediatrics, and adults were similar to what has been previously presented.

And there was evidence presented from the hemo trial as well, comments from the workgroups recognized some of the same issues surrounding the criteria for three times per week versus other, you know, pediatric, as well as the distinction between the various mechanisms of calculating this.

CO-CHAIR CROOKS: Okay. The evidence is open for discussion. I'd just like to mention there was a long list of references of additional evidence, and which in the instructions requests

a summary for each article, and that wasn't done, this long list of references. And interested as I am, I'm not going to go look up the articles to see what they say. So, again, I'd encourage the developers to follow the directions on the form. I don't know about the quality of the evidence beyond KDOQI.

DR. MADDUX: And I would just ask the developers to confirm that for the pediatric component of this there was the three or four again.

DR. DAHLERUS: Correct. So the discussion that we had for the last measure applies here as well.

CO-CHAIR CROOKS: Other discussion about the evidence? I'm sorry. Mahesh.

DR. KRISHNAN: Is there any direct evidence that a combined measure has utility? Or we're saying the fact that the two components each independently have utility means that we should confer the associative property to the combined metric, is that correct?

CO-CHAIR CROOKS: Well, I am going to presume that the reason is similar to what we heard yesterday, that for small units or units that have a small number of peds, we want to make sure the peds patients get included. But you're asking, is there evidence that this is a good measure.

DR. KRISHNAN: Yes. It wasn't just in the policy, get the policy component, the minimal cell size, the intent. Just ask about the evidence itself.

CO-CHAIR CROOKS: Right. Do the developers have any comment?

DR. DAHLERUS: So, again, we -- because the thresholds of 1.2 are based on the evidence cited in the individual measure, we would argue that this also applies to the combined measure. I think the discussion yesterday about the combined PD measure also applies here, whether there are any concerns about the values for adult somehow masking the result for the pediatric population. But, again,

that would probably be less the case given the slightly larger HD population.

DR. KRISHNAN: And I have scrolled down to validity, but was this discussed at the TEP, this combined measure?

DR. DAHLERUS: Not the combined measure. The individual measures were, though.

CO-CHAIR CROOKS: Alan?

DR. KLIGER: And just to clarify, the three or four treatments a week is true for both adults and children?

DR. DAHLERUS: No. Just children.

CO-CHAIR CROOKS: Okay. Any other comments?

DR. KRISHNAN: So, Frank and others, how did you guys rate the evidence for this?

DR. MADDUX: So having been on both 1423 and 2703, the same issues apply predominantly to the pediatric application. So for the adult portion, I think the evidence is pretty good, and that would be either high or moderate. For the pediatrics, my thought is that

this is suspect because of the same discussion we just had.

DR. KRISHNAN: How do you synthesize the two of those in one measure? Is that --

DR. MADDUX: I think the flaw in the measure, as a combined measure, has to be the same flaw as the least common denominator of the two -- of both measures.

DR. KRISHNAN: Agreed.

CO-CHAIR ANDERSON: And I guess from my perspective, if you have a separate measure for pediatrics and a separate measure for adults -- and, again, I'm not really sure what you gained by combining the measures into another measure unless you are able to harmonize the measures at the very end.

So I was struggling with all of the different measures. And then combining them into one where the evidence for the pediatrics is what we want there, certainly very strong evidence for the adults, but the pediatric evidence, even in the guidelines, was a grade C.

CO-CHAIR CROOKS: Okay. Other discussion? Measure developers.

CO-CHAIR ANDERSON: Let's get the committee --

CO-CHAIR CROOKS: Yes. I'm sorry. Lori?

MS. HARTWELL: To my understanding, combining the measures was to pick up the extra 11 kids that would not be normally measured. Is that correct? The reason for this?

DR. DAHLERUS: Correct. Yes. So more facilities and more children will be evaluated in the combined measure.

CO-CHAIR CROOKS: Mahesh?

DR. KRISHNAN: And this is just why I struggle, right? I mean, the difference between -- I get the policy objective. It makes complete sense to me. I just don't get the evidentiary basis for the construction of that. If there was an attachment, there wasn't -- there isn't strong enough evidence, if we're just evaluating -- I'm assuming we're just evaluating it based on the

evidence and not on the policy implications.

CO-CHAIR CROOKS: You dropped your card.

DR. KRISHNAN: Yes.

(Laughter.)

CO-CHAIR CROOKS: Don't say I'm not observing things up here. Okay?

Any other discussion about the evidence before we vote? Yes, Michael.

DR. SOMERS: You know, just all the discussion about the evidence and as it pertains to kids. I mean, some of it is focused on the four time a week, but, again, as I said before, I mean, where actually, if you are meeting this criteria with four time a week dialysis, you are exceeding adequacy standards.

So, you know, as a quality measure, this is ensuring that you are giving good care to those children. So, you know, I wouldn't use that necessarily to speak against this measure from an evidence standpoint.

DR. KLIGER: And while I understand

and respect what you say, Michael, I also want to point out that as written it would be a disincentive to increase the frequency of dialysis. If you have a borderline child who is not meeting the 1.2 standard single pool on three times a week, if you wanted to meet the standard you would not be able to increase the frequency.

DR. SOMERS: I think if you were to look at that five or six percent of children who are dialyzed four times a week, they are being dialyzed four times a week for volume issues. And so, you know, in terms of the amount of clearance you're getting, you're far exceeding what you ever need.

DR. MADDUX: If I could comment, too, Michael, the way I look at this is we have a particular set of measures here, and we have opportunities through the process to get it right. Rather, in my opinion, get it right or as right as we can understand it, than to try to say it's right enough, because I do think as we went through previously there are multiple ways in

which to not diminish the opportunity to look at the pediatric population and those on dialysis, and yet also not have some of the issues with the particular measures.

CO-CHAIR CROOKS: Okay. It looks like we're ready to vote on the evidence.

MS. OGUNGBEMI: The committee is now voting on evidence for Measure 2703. The options are one, high; two, moderate; three, low; and, four, insufficient. Voting is open.

MS. BAL: We're contacting IT about that, but that screen is available.

CO-CHAIR CROOKS: So unlike the last one where we voted no, this is a gray zone. It's above 40 percent, with a one or a two.

MS. OGUNGBEMI: The results are zero votes high, 10 votes moderate, nine votes low, and four votes insufficient. So the measure is gray zone, Measure 1423.

CO-CHAIR CROOKS: Thank you. Sorry I got ahead of you. Just doing numbers today.

Okay. So I think we don't need to

vote for exception in this case, or maybe we do just for insurance. I don't know.

MS. BAL: No.

CO-CHAIR CROOKS: No.

MS. BAL: We would move forward as if this measure had been accepted.

CO-CHAIR CROOKS: Okay. So we'll continue this gray zone measure and go on to the performance gap.

DR. MADDUX: So the gap -- the discussion of the gap recognized the data for both the CROWNWeb and the Medicare claims data, and the performance of the elements, the pediatric element and the adult element. I don't think there was substantial disparity recognition there.

And the comments of this measure surrounded the similar comments that we have had related to the predominant driver of this aggregate measure being adults, because of the number of patients, questioning whether it has topped out, knowing that independently the adult

looked like there was not much of a performance gap, and there was a mild to moderate performance gap in the limited number of patients tested in the pediatric population.

CO-CHAIR ANDERSON: No further comments, other than the risk adjustment. There was not significant -- and it does, in the 2015 QIP, it is at -- Kt/V is at 98 percent.

So, again, there is little room for performance improvement in the adult population. And I think that's one of the difficulties in combining the measure is that you have -- the adult population really is pretty well topped out and has shown increased improvement.

CO-CHAIR CROOKS: All right. Other comments about the performance gap?

DR. KRISHNAN: Frank, is this the exact -- Connie, is this the exact same number for the performance gap from the adult measure? It seems like it, right? The 93.5 percent?

CO-CHAIR ANDERSON: In the 2015 QIP, it is 98 percent.

DR. KRISHNAN: No, I understand. But in the performance gap for this measure, the combined measure compared to just the adult measure --

CO-CHAIR ANDERSON: Oh, yes.

DR. KRISHNAN: -- it is the exact same number?

DR. MADDUX: So, Mahesh, I did not interpret this as being the aggregate, at 93-1/2, that as being the pediatric. I may be wrong, but --

DR. KRISHNAN: It is just the adult, again, so that is the exact same number. So you don't have a gap for the combined?

CO-CHAIR ANDERSON: I think the 93-point is the 2013 percentages and is the gap analysis combined.

CO-CHAIR CROOKS: Yes. Can you clarify what the gap for this measure was when it was determined?

DR. DAHLERUS: So in the section on meaningful differences, we do show that about 15

percent of facilities are performing worse than expected, and that would -- and so we provide that in the meaningful difference analysis.

CO-CHAIR CROOKS: Right.

DR. DAHLERUS: And just to clarify, this combined measure is not implemented in the QIP.

CO-CHAIR CROOKS: Do you have a mean performance and a percentage?

DR. DAHLERUS: Ninety-three percent.

CO-CHAIR CROOKS: Yes. That's what I -- okay. This measure is at 93 percent. They can calculate outliers, you know, statistically, but the mean is 93.5 percent.

Okay. Other discussion on the gap? Okay. Let's vote.

MS. OGUNGBEMI: The committee is now voting on performance gap for Measure 14 -- or 2703, pardon me. The options are one, high; two, moderate; three, low; and, four, insufficient. The voting is open.

Results are one vote high, eight vote

moderate, 11 votes low, and three votes insufficient. And the measure does not pass on performance gap. It's Measure 2703.

CO-CHAIR CROOKS: Okay. So that puts it in the category of possibly being eligible for reserve status. So we will continue.

MS. BAL: If the committee would like to consider it for reserve, we can do a hand vote. But it does not have to be considered in that manner. So it would be more, do you feel that this should be a candidate? If not, this measure will fall here, and we will move on to the next one.

DR. MADDUX: So I have one question. A measure that has never actually been used can be put in reserve? It's like being the designated hitter forever.

(Laughter.)

CO-CHAIR CROOKS: Sarah? Can you help us with this question? We have even managed to baffle Sarah. All right. Very good.

DR. ANDRESS: Okay. Sure. So I think

from a certain standpoint it makes some sense. And we talked about the adult measure going -- as possibly going into reserve status as a candidate for that. You know, and the reasons, of course, for that measure also apply here because the vast majority of the denominator in the combined measure is driven by that measure. So it makes a certain degree of sense.

And I think the classification for why it would -- if you think of it as an ultimate potential replacement measure, then certainly that makes sense, because the same reason for having the measure in place still stands.

You know, there is the additional concern of course of the specifications for the pediatric population that are combined within the measure, but I think that is a separate issue from whether or not it should be considered for reserve status, you know, of course pending the -- - pending our efforts to kind of resolve and get back to you the specification issues.

I mean, I think, you know, the purpose

ultimately of this measure is to take on the -- you know, to generally take on the same role as the measure that is going into reserve status, and so I think it makes sense to consider it for reserve status on that basis.

I don't know procedurally if NQF would allow for that, but I think it's -- you know, I mean, in some ways, you know, we propose it here as a distinct measure, or we present it here as a distinct measure because, you know, there are the measures implemented, and we wanted to ensure that we retained a continuity of endorsement status if the combined measure was not something that the panel was willing to consider.

But it can also be considered a modification of the existing measure to try to address at a particular policy need. At the end of the day, it is addressing -- it is assessing the adequacy of dialysis received by these populations.

You know, if that is something that we consider to be an important measure of quality

that goes forward, even given the high level of performance in the overall population, then I think the reserve status -- I think it makes sense. And, again, you know, it's -- whether or not NQF can support that, I -- you know, I leave that to Sarah to respond.

MS. SAMSPEL: So I'll ask Michael to make his comment, and we are still gathering our responses.

DR. FISCHER: If I could just a quick -- I mean, I looked at the -- because we looked at this yesterday, the NQF documents, and going to Franklin's point it says this is for -- talking about retaining endorsement.

And, specifically, there is a line in here that says, "Would not grant an active endorsement status for a measure that has not been used." This measure has not been used before. So I guess, if I understand your own documentation correctly, kind of -- this isn't possible in this situation with a composite measure that has not been promulgated and used

previously.

MS. SAMSPEL: So I think technically we could argue both ways. This is unprecedent-setting. You know, I really -- we don't have a precedent to go back to on this one. But we have the issue of a component of this measure is an endorsed NQF measure that you earlier voted reserve status.

The other portion of this measure, another component of this measure, you know, is in gray zone because that was the one that -- you know, and thus has not -- is not implemented or not endorsed.

And so -- and really going back to what, you know, you just pulled out as well, Michael, is, you know, I -- in light of the fact that the second component of the measure actually has to come back to revote as well, I mean, what I really would do is not deal with reserve status right now at this meeting until we can deal with the rest of the measure issues through the public comment period and through post-public comment

voting on the second half of the measure.

CO-CHAIR CROOKS: Well, I don't know about you guys, but I'm getting dizzy. So then at this point, then, we'll stop considering this measure, as I understand it, and move to the next.

Okay. That cleared my head.

2705 is the next measure. This is, as we know, a CMS measure combining other measures. So if you are ready to -- are you ready to go ahead?

DR. DAHLERUS: Yes.

CO-CHAIR CROOKS: Okay.

DR. DAHLERUS: Yes, we're ready. Are we ready? Okay. So, again, in the interest of time, just to clarify. So this is the overall combined dialysis adequacy measure. So it's a combination of each of the individual measures that we have discussed in the past two days.

And so this is -- and, again, the intent here was to include as many patients and facilities and the assessment of adequate

dialysis. So it's a combination of the respective pediatric hemodialysis and peritoneal dialysis adequacy measures, and then the respective adult HD and PD measures as well.

This is -- the combined measure is a new measure, and it has not been implemented in Dialysis Facility Compare or QIP. And, again, the stipulation about the upper threshold is something that will also be revised for this overall combined measure.

And in terms of the evidence that we summarize in the evidence form, the evidence draws on specific minimum thresholds for each of the distinct populations.

CO-CHAIR CROOKS: All right. Our primary reviews are Alan and Elizabeth. Who would like to kick it off?

DR. KLIGER: Yes. So there are two facts about this measure that are I think different than the others. It obviously has the same issue that we have just spent the last torturous two hours talking about. But, in

addition, if you look carefully at what they have said, the denominator statement says for adult hemodialysis patients receiving dialysis, excluding those who receive dialysis less than three or greater than four times a week. So this does include indeed adults getting four times a week dialysis.

The other thing to note that is somewhat different is unless I misread this, it does not exclude home dialysis patients. And if it does not, then there have been, as we all know, an increasing number of adults and kids, but mostly adults, who are dialyzing at home with much higher frequencies, five, six times a week treatments using equipment that facilities that.

So the evidence issues are the same here, except that it does include adults on variable frequency of dialysis, and includes a larger -- you know, larger than a minimal number of adults for whom that really is an issue.

The evidence is the same evidence, and I hate to go through it again. It really is the

same, and I think the fatal flaw is the same one.

MS. EVANS: I have nothing else to add, but I would like to ask, could we not, to save time, put this on hold right now? I mean, it's the same issue that we had before, rather than going through all of this again?

DR. DAHLERUS: So the denominator exclusions state that adult HD patients receiving less than three or greater than four -- four or greater times weekly dialysis are excluded. So they are not included in the denominator.

DR. KLIGER: Correct. That's what I said. Greater than four is excluded; four is included, not excluded. It's not greater than or equal to four. The specification says greater than four.

DR. DAHLERUS: So we apologize. That is something that was not clearly stated in the denominator. The statement for the denominator does limit it to thrice weekly.

CO-CHAIR CROOKS: Yes. I agree with Alan's reading. It doesn't say excluding four or

greater; it says excluding greater than four.

DR. KLIGER: Correct.

CO-CHAIR ANDERSON: Okay. I think what has happened is there is a discrepancy in what they wrote, because in the denominator statement they do say thrice weekly for the adults. But in the denominator exclusion, which is where the discrepancy is, it says greater than four times a week. So you can't -- they're inconsistent.

CO-CHAIR CROOKS: So we've had a request to shorten the process, if possible. I'm not sure how we can -- does the staff have suggestions on how we --

MS. BAL: I mean, we can just advise the committee not to rehash issues that have already been rehashed and just speed to the vote. But we do need to vote on this measure.

CO-CHAIR CROOKS: Okay. Thank you.

So is there further discussion on the evidence before we vote? Okay. Let's vote.

MS. OGUNGBEMI: The committee is now

voting on Measure 2705, evidence. The options are one, high; two, moderate; three, low; and, four, insufficient. Voting is open.

Results are zero votes high, four votes moderate, 12 votes low, and six votes insufficient. Measure 2705 does not pass on evidence.

CO-CHAIR CROOKS: So that brings us back to the possibility of an exception. Does anyone want to propose that and make the case for that? Okay.

So that being the case, I think we can end discussion of this measure at this point. Okay, Elizabeth?

All right. So I have good news for you. It's break time. We are going to come back, and then we are going to jump into the related competing measures for all of the Kt/V measures. And we'll be moving on to --

CO-CHAIR ANDERSON: We are not going to be doing the related and competing --

CO-CHAIR CROOKS: Oh, we're not. Oh.

Because?

MS. SAMSPEL: So the issue, if you'll -- the issue with related and competing measures is that they are only discussed for the measures that have been recommended for endorsement. So we have measures that both failed and went into gray zone or consensus has not been reached.

So we will be taking that offline, and the related and competing discussion will happen post comment, post public comment, and those calls. So when we come back, we will be jumping into the next set of measures.

CO-CHAIR CROOKS: Thank you for clarifying.

Okay. So our break is scheduled for 15 minutes, which would get us back at 10:05.

Thank you.

(Whereupon, the above-entitled matter went off the record at 9:50 a.m. and resumed at 10:04 a.m.)

CO-CHAIR ANDERSON: We're going to change the order of the agenda a little bit

because of the developers that are available to us. So, instead of starting with 1424, we're going to start with 1667, then do 1660, and then go back to 1424 and start there. So, Dale and Amy. 1667.

MS. BECKRICH: So, we should have Barbara Fivush on the line, and she's going to introduce 1667.

DR. FIVUSH: Thank you. This is Barbara Fivush. Can everybody hear me?

CO-CHAIR CROOKS: Yes, Barbara.

DR. FIVUSH: All right, thank you. So, 1667 is a Pediatric ESRD Patients Receiving Dialysis Hemoglobin less than 10 measure. It's the percentage of calendar months within a 12-month period during which patients aged 17 years and younger with a diagnosis of ESRD receiving hemodialysis or peritoneal dialysis have a hemoglobin level less than 10.

The rationale for this measure is that in pediatrics, as you all know, anemia is a major complication of ESRD in children. And the anemia

is associated with cardiovascular dysfunction, cardiomyopathy, and death. Correction of anemia in children has been shown to improve cardiac dysfunction, exercise tolerance, and reduce left ventricular hypertrophy.

We also strongly weigh the potential quality of life benefits that correction of anemia will bring, and the available literature in adults should be enriched by consideration of quality of life measures that are critical to children, including neurocognitive development, school attendance, exercise capacity, and family support.

The measure has recently been reclassified as an intermediate outcomes measure, and it is a physician-level measure. Currently, it is being measured in PQRS, and included in the RPA Kidney Quality Improvement Registry for 2015.

There is some strong evidence for this measure. Their KDOQI Clinical Practice Guidelines, 21.1 in particular, which say the lower limit of hemoglobin which is fully

applicable to children in patients with CKD, the hemoglobin level should be 11 or greater, and that's a very strong recommendation.

So, Pediatrics feels strongly. This has been approved in the past but this measure should be reapproved, and agree with the reclassification, as you believe that measurement of hemoglobin is critical to insure the best care for children with ESRD on peritoneal and hemodialysis.

CO-CHAIR CROOKS: Okay, thank you. The discussants are Bobbi Wager and Frederick Kaskel. Have you decided who's going to be on first?

MS. WAGER: Yes, this is Bobbi. I will go ahead and start. Thank you, Barbara. She did an excellent job. The evidence, as she pointed out, the KDOQI Guidelines, and the data also comes from different patient registries and cohort studies demonstrating the adverse effects of anemia. We all know, as she stated, cardiovascular disease, et cetera.

Myra, do you have anything else that

--- oh, I'm sorry, it's Dr. Kaskel.

DR. KASKEL: It's a very important measure, obviously, as Barbara has summarized very nicely.

CO-CHAIR CROOKS: I have a comment, that the measure calls for hemoglobin less than 10. The guidelines state the target should be between 11 and 12. And is there justification in the evidence why that particular number is targeted? Michael.

DR. SOMERS: So, there is evidence in children with CKD and ESRD that hemoglobin values less than 10 are associated with higher rates of hospitalization, higher rates of mortality, as well as the quality of life standards, and the cardiac effects that have been discussed previously.

CO-CHAIR CROOKS: Okay, that's --- that evidence wasn't, I don't think, in the submission, but duly noted. So, that as a pediatric nephrologist you think the number 10 is valid, or is a good choice?

DR. KASKEL: I believe it's a good choice and, if anything, we tend to even think about higher levels depending on the age and conditions. Neurocognition are very difficult area to measure and quantify, needs to come into this picture, as well, at certain periods of development. And there may be racial disparity issues.

DR. SOMERS: And even though that's --- the evidence supporting less than 10 isn't provided specifically with this measure, but it's going to come up in 1424 when that evidence was actually given.

CO-CHAIR CROOKS: Very good.

MS. WAGER: Peter, can I make a comment? On our conference call, I believe the developer was asked why below 10, and I believe they said something about they wanted a floor level, or the bottom level. That's why it was below 10.

CO-CHAIR CROOKS: Right. I was just wondering if there's also data or medical

evidence that, you know, makes that appear to be the right target. Lorien, you have your card up.

DR. DALRYMPLE: Can I just clarify from the work group how they graded the evidence using the NQF Criteria? Was this a clinical practice recommendation from KDOQI? That's just what I wrote in my notes, but I'd like to clarify how the evidence rated by NQF criterion?

CO-CHAIR CROOKS: Do you remember what grade level it was assigned? I can ---

MS. SAMPSEL: I don't necessarily that in each of the work groups that the Committee on the work group call that they went through the official process. I mean, what I would ask though is that, you know ---

DR. DALRYMPLE: What are their thoughts after discussion ---

MS. SAMPSEL: The Committee members who looked closely at this, did you have any thoughts on that?

DR. DALRYMPLE: Yes, did you think this was perhaps insufficient because of pediatrics

and the challenges around it, or was there a higher grade? And we've talked about insufficient with exception because we think it's important, but I just was hoping to get those who looked at in depth for their sense of where this fell on evidence.

DR. SOMERS: In terms of the level being less than 10, I do think there is specific pediatric evidence that supports the adverse effects of your hemoglobin being less than 10, and that mark is supported by pediatric evidence.

CO-CHAIR CROOKS: According to the submission, the grade from the KDOQI process was moderately strong, which is their second-level grade.

Let's see. I think Michael Fischer was next. That was accidental card turning. Dr. Wagner.

DR. WAGNER: Thank you. Just so I understand the process; if the evidence is not in this particular submission, even though it is in another submission, does --- are we allowed to

then consider that, or are we obliged to follow what has been in this submission?

CO-CHAIR CROOKS: I would say that when it comes to evidence, in our past discussions, Sarah, we heard that if the Committee is aware of other evidence that bears on the measure, that it's okay to consider it at this level. That may not be true when it comes to other parts of the submission, but I think if you're aware of evidence that supports it, you can use that in making your reasoned judgment. Okay? Mahesh.

MS. HARTWELL: I had my ---

CO-CHAIR CROOKS: Oh. Well, I'll call on Mahesh first, because I'm the Chair. I can do that. No, I'm sorry. I didn't see Lori.

MS. HARTWELL: I think you're ignoring me.

CO-CHAIR CROOKS: You're just --- when I have my glasses --- okay, Lori.

MS. HARTWELL: I just had a quick question because it talks about transfusion avoidance, but also it doesn't make mention of,

you know, when children get blood transfusions, it's more difficult to transplant. And I just wanted to have any comments from the Committee about that, because it's a very serious issue for children if they get transfused.

CO-CHAIR CROOKS: What about the impact on blood transfusion, pro or con, does it work against that?

DR. KASKEL: We would really worry about having to transfuse with the attendant risks of infection, but also the need to keep them less sensitized for transplant for sure. We would want to avoid that at all costs.

CO-CHAIR CROOKS: Does the measure push from --- end up causing more transfusions because of that floor, than maybe you would do otherwise?

DR. KASKEL: Well, this is a big issue to consider down the line. If the Pediatric Centers caring for the children would want to keep the level, if anything that would be the minimum level here at this floor, but be higher to avoid the need for transfusion.

CO-CHAIR CROOKS: So, if you have a ESA-resistant child and the hemoglobin is 9, you're going to have to transfuse him, probably.

DR. KASKEL: In rare instances, if you can't correct the resistance and they're symptomatic, you would.

CO-CHAIR CROOKS: So you try everything you can to avoid it, obviously.

DR. KASKEL: Right.

CO-CHAIR CROOKS: Yes, Josh?

DR. ZARITSKY: And, ironically, it's usually an upper limit that leads to this type of thing because no one is going to transfuse a patient to make a recommendation. Do you know what I mean to say? Oh, my God, you know, I'm not going to make my, you know --- whatever you want to say, and then transfuse the patient. Whereas, if you look at probably the effects of the upper limit, that might actually have a little bit more of a significance, or the range that you're trying to keep the patient in rather than the lower limit. I hope that helps.

CO-CHAIR CROOKS: Rick.

DR. KASKEL: I just want to mention that this area of anemia in CKD in children is being actively studied through some cohort studies. And the CKiD Study, Chronic Kidney Disease in Children, is one of the main ones. This is for children with CKD Stage 2 and 3, and anemia is one of the major comorbidities that's being evaluated in terms of progression of kidney failure, neurocognition, growth, and cardiovascular risks, so it's an active, ongoing collaboration of most of the large pediatric centers across the country, something like 600. So, it's ---

CO-CHAIR CROOKS: Okay, good to know.

DR. KASKEL: --- a work in progress that you need to know that from our standpoint, the information that we're getting about appropriate targets at least in CKD is accurate.

CO-CHAIR CROOKS: Okay. Mahesh.

DR. KRISHNAN: I just had a question for the pediatricians. Anemia management in

adults has clearly got a seismic shift in the last four or five years. The guidelines that are referred predate that. I don't know enough to know enough about how much that's impacted the pediatric guidance. Could you guys help us with that?

DR. KASKEL: This is a concern because when the black box warning came out, you know, from the two studies; obviously, not including children, it --- we were concerned that it would set a level lower for our field. And there are no randomized controlled trials of target --- adequate target levels in pediatric nephrology for the hemoglobin for ESA, that's one. And the rate of rise of hemoglobin is another consideration which in the adult studies was one of the major factors for the cardiovascular morbidity. We have no idea in children whether the rate of rise of hemoglobin is a risk factor, so there's a lot of unanswered questions about this. And the black box warning did siphon down to us, and knowing that we may need a higher

hemoglobin in some children for growth. Okay, so it's a concern.

DR. KRISHNAN: Is it accurate to say that there wasn't really much of a --- you guys didn't extrapolate anything from the adult studies to kids with all the shifting?

DR. SOMERS: In terms of the higher limit, I mean, there's absolutely no data whatsoever in children that having a hemoglobin above 12 results in any sort of adverse outcome. And unlike the adult data, there is specific data showing that hemoglobin less than 10 results in adverse outcomes in children.

CO-CHAIR CROOKS: Okay.

DR. SOMERS: Specifically, that level of 10.

CO-CHAIR ANDERSON: I'm trying to keep the discussion to what we need in order to vote. Dr. Wagner.

DR. WAGNER: Quick question. So, are the data specific to ESA treated patients or all comers?

DR. SOMERS: The data that --- less than 10 dealt with kids with CKD. I'm not sure what proportion of them were receiving ESAs, but as common practice, since in general pediatrics there's evidence that any normally healthy child having anemia results in all kinds of adverse outcomes that pediatric nephrologists tend to start treating when we start to see anemia.

CO-CHAIR CROOKS: Okay. I don't see any other cards up, so let's vote on the evidence.

MS. OGUNGBEMI: The Committee is now voting on Measure 1667's evidence. Options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are five votes high, 13 votes moderate, three votes low, and one vote insufficient. Measure 1667 passes on evidence.

CO-CHAIR ANDERSON: Okay. Bobbi, on the gap.

MS. WAGER: The performance gap included, for instance, racial disparity because it's such a small scale, a small size sample of patients that it was difficult to differ

treatments with the gender, the patient's gender, age, and race. 2008 CMS PQRI data is showing about a third of the patients not meeting the goal, and 2010 Elab data provided showing that 20 percent of pediatric patients have hemoglobin values less than 10. And that proportion increased a bit over 2009. But, again, the performance does not include really a thorough investigation due to the small sample size.

CO-CHAIR CROOKS: Rick, any other comments on the gap?

DR. KASKEL: No, there is clearly a percentage of patients not meeting what we think is a target. That is somewhat of a concern to our community, and even in the CKiD data set we found a surprising number of patients in Stage 2 or 3 who were anemic and were not getting effective treatment, so it's a concern.

CO-CHAIR ANDERSON: Okay, other considerations on the performance gap? Lori.

MS. HARTWELL: The data says 2008. What year was the black box warning; so this data is

prior to the black box warning, so it might be different. Is that correct? I was just saying the gap might be higher.

CO-CHAIR ANDERSON: If I'm reading this right it says --- and, Bobbi, I'm probably --- this may be repetition, but the 50th percentile, 66.23, does that mean that's two-thirds of kids are below 10, or is that for the 10 to 20 --- maybe that's for the 10 to 20 range. Yes, please.

MS. SINGER: There was data from the International Pediatric Peritoneal Dialysis Network Registry of 1,394 pediatric patients showing that 25 percent of patients had hemoglobin levels below the target. That's in the evidence.

DR. ZARITSKY: You have to --- I don't want to argue against this, but the IPPM, you know, International ---

MS. SINGER: Right. It is international.

DR. ZARITSKY: They don't get --- they

don't have Epogen available, and there's a cost.

CO-CHAIR CROOKS: So, whether it's 15 percent or 16 percent, I guess there's a large performance gap. Okay. Any other discussion on the performance gap? I see voting machines in the air. Let's do it.

MS. OGUNGBEMI: The Committee is now voting on the performance gap for Measure 1667. Options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are for performance gap, 10 votes high, 13 votes moderate, zero votes low, and zero votes insufficient. Measure 1667 passes on performance gap.

CO-CHAIR CROOKS: Okay. Next up, specifications and reliability.

MS. WAGER: Okay. Reliability, measure was tested at the critical data element level using inter-rater reliability for medical culture rater for the measure, where the patients were randomly selected from visits. For ESRD, they used four nephrology practices in the Midwest,

the West, the East, and Southern. The number of patients seen per month was averaged at 240 to 2,800. Sample size per physician organization ranged from 24 to 30 for a total of 169 patients on PD or hemo.

The data was collected from medical records, data abstraction completed for multiple patient visits per patient for a total of 2,012 patient visits. Data abstraction was done in 2008, and testing results indicate from that there's a high reliability.

CO-CHAIR CROOKS: Okay. Rick, any other comments on ---

DR. KASKEL: No, that covers it.

CO-CHAIR CROOKS: The source of the data is from CROWNWeb, or it's from some other ---

MS. SINGER: Chart extractions.

CO-CHAIR CROOKS: For the testing.

MS. SINGER: Yes.

CO-CHAIR CROOKS: But, in general?

MS. SINGER: PQRS.

CO-CHAIR CROOKS: PQRS. Thank you. Okay. Any other comments on specs reliability?

DR. DALRYMPLE: Can I ask a question of either the ---

CO-CHAIR CROOKS: Oh, Lorien, sorry.

DR. DALRYMPLE: --- developers? Oh, sorry, I thought you were ---

It was --- I think when I looked at this, the reliability testing seemed to be the same as the adult measure. So, can we get clarity as whether the reliability testing occurred in a pediatric population? And then I think the kappa I saw was for a plan of care for anemia measure, and is that the same as a hemoglobin less than 10? And maybe I'm missing the hemoglobin less than 10 kappa, but I was just hoping for clarity on the reliability testing.

MS. SINGER: Yes, we did not test this in pediatric practices, so it's the same testing that was done for the other measures.

CO-CHAIR CROOKS: Is there any reason to think the reliability testing would be

different if you'd done it on pediatric patients?

Why is the sky blue? Oh, we know that ---

DR. DALRYMPLE: Can we discuss it as a Committee whether that changes things for anyone, the lack of pediatric sampling, or the kappa is on a plan of care for anemia, as opposed to hemoglobin cutoffs, which is what I think the focus is in this one?

CO-CHAIR CROOKS: I agree, it does raise some questions. Michael.

DR. FISCHER: Quite simply, the reliability testing is supposed to be on the measure proposed. If it's not, I'm not sure how it's germane to the application.

CO-CHAIR CROOKS: Other comments?

DR. KASKEL: One of the deficiencies already mentioned was the fewer numbers of patients and inability to get any real signals regarding gender and race, and certain periods of development like pre-pubertal/post-pubertal. These are all lacking in the analyses. We don't

have enough patients entered.

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

MS. OGUNGBEMI: The Committee is now voting on reliability for Measure 1667. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results for reliability on Measure 1667 are one vote high, eight vote moderate, four votes low, and 10 votes insufficient. The measure fails on reliability. It's Measure 1667.

CO-CHAIR CROOKS: That's a critical element, so do we stop at this point, or do we -- --

So, we would like to offer the Committee the opportunity to provide additional comments to the developer.

DR. KASKEL: I'm concerned about this now, so I'd like to have a further discussion about the reliability in detail.

CO-CHAIR CROOKS: So, are you arguing that the reliability was --- testing was

suitable?

DR. KASKEL: My comments about gaps related to need for further studies, just like target hemoglobin and all the factors that go into the distribution, racial, gender. But the reliability of this assessment, I don't believe was insufficient from the data source. That's a concern.

CO-CHAIR CROOKS: I need to make sure I'm understanding you; that you think the reliability testing is sufficient, or it is not sufficient?

DR. KASKEL: For this measure, it's sufficient.

CO-CHAIR CROOKS: You think it is sufficient. I think to summarize the opinion I'm sensing from the Committee, that when you don't test it on the target audience and show that your bullet is hitting in the same place every time because you're not even looking at that population, it's not being judged reliable. That's my sense of it. Let's start with --- who's

up first? Michael.

DR. SOMERS: It may not have been tested on a target population, but I think it would --- I'm not sure why doing record abstraction looking at the same data element in adults would be different in a pediatric population at these facilities, as well. So, I'm not sure the outcome of reliability testing, if there were a pediatric cohort that could have been tested, would have been any different than, you know, this measure being tested in adults, because the process is the same of looking at something ---

CO-CHAIR CROOKS: Right, we hear you. Yes.

DR. SOMERS: Yes.

CO-CHAIR CROOKS: Okay. Lori.

MS. HARTWELL: I just want to clarify, we're just talking about testing hemoglobin. Correct? That's the test. I mean, hemoglobin --

CO-CHAIR CROOKS: No, the reliability test is that the data elements, whoever finds

them, or whatever method they're found, they're going to come out the same.

MS. HARTWELL: So, can you explain that to me just a little bit relating to hemoglobin? I mean, is it the method --- I'm trying to understand this one, because I'm like I don't ---

CO-CHAIR CROOKS: Yes, it's beyond the scope of my duties as Chair to explain reliability beyond what I can, but the notion is that the data elements are --- however they're brought in, that they are accurate, that the number --- it doesn't matter how high or how low the number is, it's that it's accurate time after time it's being compared.

MS. HARTWELL: Just so I can understand the process, so the data elements, if it's brought in an adult, they measure the hemoglobin, it comes in and it's the fact that it wasn't measured on the pediatric patient the same number is making it not reliable? I'm just trying to understand this. Is that ---

CO-CHAIR CROOKS: Well, we don't know if it's reliable in the pediatric patients. We don't know if they're missing values. We don't know that inter-rater reliability would have been good. We don't know, because they didn't do it. And that's causing us some --- giving us pause. It's very important that if we endorse a measure, that it's reliable and valid.

MS. HARTWELL: Okay.

CO-CHAIR CROOKS: And that's part of the NQF mission. Measures that matter, if they're not reliable and valid, why should we be putting a lot of health care resources into making them better when we don't know that they work?

DR. FIVUSH: Can I speak as the measure developer?

CO-CHAIR CROOKS: You may have a brief comment.

DR. FIVUSH: I would strongly support Michael's comments on this. There would not be a reason to believe that this would be different in a pediatric or adult population, not the level of

hemoglobin, but the reliability of being able to obtain the hemoglobin value that is accurate from chart abstraction. I can't --- I think that his points are very valid, and there would not be a particular reason, or any reason to assume that this would be different if it was tested in the pediatric population by chart abstraction.

CO-CHAIR CROOKS: Okay, thank you. We have other people who would like to speak. Michael.

DR. FISCHER: I don't want to be too rigid. I guess, my vote was informed by following this algorithm, but perhaps I went to the boxes improperly. But given there's some concern about reliability, as Lori mentioned, perhaps it would be useful ---

CO-CHAIR CROOKS: Yes.

DR. FISCHER: --- for us to --- which I think is what you were going to do, anyway.

CO-CHAIR CROOKS: Can you step us through it really quickly, the recommended way to consider it?

DR. FISCHER: Yes, so --- and I'll be candid. I wasn't part of the working group for this measure, so I'm basing how I went through on the preceding comments, but it's Algorithm 2, I think that's page 12.

So, are submitted specifications precise, unambiguous, and complete, they can be consistently implemented? Yes. So, I went to number 2. Was empirical reliability testing conducted using statistical tests that the measure specified? No, because once again, that's where I thought about the population thing, but I don't --- once again, just trying not to make assumptions. Was empirical validity testing of patient level data conducted? When I looked quickly it was no, but I know we haven't talked about validity yet, and that's how I wound up with my vote. But, once again, I'm happy as others look at this, I can be corrected, but I was just going through this as to the NQF process.

CO-CHAIR CROOKS: Okay, thank you.

Lorien.

DR. DALRYMPLE: Can I ask the developers a point for clarity? So, even if we accept that perhaps there are no differences in the adult --- in the pediatric population, the one thing I want to make sure I understand is the kappa that's provided for plan of care of anemia the same as this measure as specified percentage with a hemoglobin less than 10, or is that actually a different measure?

MS. SINGER: The plan of care was part of the original measure. It's not part of this current measure. And the current endorsed measure does not have a plan of care component, either. So, I guess to clarify the reliability testing issue, though, the process for testing the measure would be the same in the pediatric population, and I just want to emphasize that point. And the numbers of patients are so small that the cost involved in the testing was placed in the larger population of patients so that we could get larger numbers. But the process is the

same for testing.

DR. DALRYMPLE: But this kappa does not apply to the measure as specified, even if we negate the pediatric issue. Is that the correct understanding?

MS. BECKRICH: So, the measure, I mean, as it was tested would include the hemoglobin element as specified in the measure. And, in addition, it included a plan of care element which was in the original, I think, 2007 version of this measure, but that measure --- that element of the measure was dropped when it was brought before the NQF Committee last time, I believe 2012. So, the current endorsed measure does not include the plan of care.

CO-CHAIR CROOKS: Frank.

DR. MADDUX: So, just to --- you had asked, Peter, if we'd make comments to the developers given the result of the vote, and it strikes me that what we're really talking about here is for reliability testing using a proxy. And you're using a proxy to --- so, part of the

defense I think might be more than just elucidating the data of the proxy, but why the proxy can be used in a valid way for reliability testing on this particular measure, because a lot of the kinds of things you're been talking. I think if that path was laid out, there probably would have been a different result.

CO-CHAIR CROOKS: I think I'm sensing, though, a request to re-vote on the specs and reliability. Are there any objections to the Committee --- from the Committee to re-vote? Okay, let's re-vote.

MS. OGUNGBEMI: The Committee is now voting on reliability for Measure 1667. Options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are one vote high, 15 votes moderate, one vote low, and six votes insufficient. The measure passes on reliability. That's Measure 1667.

CO-CHAIR CROOKS: Should we make it two out of three? No? Okay, we'll go on. Validity.

DR. KASKEL: So, for the validity the

comments from the group were that incorrect entry of data and codes, not following harmonized reporting at specific time points may have affected some results. There's missing data. The test sample is adequate, however, for widespread implementation; however, normative adult data should not be used for comparison to pediatric age, gender, and race specifications.

Medical record abstraction data from over 2,000 patient visits showed appropriate reliability for medical record abstraction. So, moving forward with EHR, this can only improve, especially with CROWNWeb. The panel thought this was high validity from the TEP.

CO-CHAIR CROOKS: Elizabeth, any other ---

Bobbi, sorry, Bobbi.

MS. WAGER: No, no more comments.

CO-CHAIR CROOKS: You're so far away. Okay, other --- so, it was a TEP --- basically, it was based on TEP recommendation, the validity argument. Any other discussion on validity? Okay,

let's vote.

MS. OGUNGBEMI: The Committee is now voting on validity for Measure 1667. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are three votes high, 18 votes moderate, zero low, and one vote insufficient. Measure 1667 passes on validity.

CO-CHAIR CROOKS: Very good. So, we're up to feasibility. Bobbi or Rick?

DR. KASKEL: So, for the feasibility, again this was electronic sources, data collection, and care processing, so the electronic reporting needs to be --- it was felt needs to be more complete and uniform across sites. Analyses of data needs to target specific factors, such as, again, age, gender, and race, and the normative --- same comments about normative adult population data should not be used to assess performance in the pediatric population. And, again, this will improve with electronic medical record, or electronic health

records.

CO-CHAIR CROOKS: Bobbi, any other thoughts; feasibility?

MS. WAGER: No comment. He said it.

CO-CHAIR CROOKS: And, again, this is PQRS data, so it's submitted by physician practices?

MS. SINGER: Individual physician providers.

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

MS. OGUNGBEMI: The Committee is now voting on feasibility for Measure 1667. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are for feasibility 10 votes high, 13 votes moderate, zero votes low, and zero insufficient. Measure 1667 passes on feasibility.

CO-CHAIR CROOKS: Okay. Bobbi and Rick, is it in use, and is it usable?

DR. KASKEL: Sorry. This is a public database and accessible because of that. And it's

--- the data can be used to determine anemia management and records of deviation from expected norms can be used to investigate specific etiologies of inadequate response to anemia treatment, which is a good thing to have. And the benefits, however, at present outweigh any unintended consequences. We did also recommend that data provided more recent than 2010 would be useful as we move forward; that it was transparent, as well.

CO-CHAIR CROOKS: Bobbi, anything else? I think they stated in the submission that there is evidence for improvement that the curve is shifting to the right. Is that a result of their measure? Maybe. Okay. Other discussion on usability and use? Okay, I think we're ready to vote.

MS. OGUNGBEMI: The Committee is now voting on usability and use for Measure 1667. Options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are 14 votes high, nine votes moderate, zero votes low,

and zero votes insufficient. Measure 1667 passes on usability and use.

CO-CHAIR CROOKS: Okay. Before the final vote to submit it for endorsement, open it up for general comments. Just to summarize from the pediatric nephrologist's point of view, maybe if I can --- you can tell me if I'm right. You feel this is an important measure. It's the only measure --- NQF measure relating to anemia management in children, and you think it's a good measure.

DR. KASKEL: I think it's very important as we move forward. And I think we're on the verge of having more information regarding the etiology of the anemia in certain groups, such as with disparities. We have some information on hepcidin and racial disparities in hepcidin and how that may affect anemia, and cardiovascular and neurocognitive outcomes, and growth.

DR. SOMERS: And I think the fact that it's very much data based with the target level

of 10, is very important to the pediatric community.

CO-CHAIR CROOKS: Okay. Any other comments before vote? Okay, let's do it.

MS. OGUNGBEMI: The Committee is now voting for Measure 1667's overall suitability for endorsement. The options are 1 yes, 2 no, and voting is open. The results are 20 votes yes, and three votes no. The measure passes for meeting NQF criteria for endorsement.

CO-CHAIR CROOKS: Well, that was a squeaker, but the process worked. Thank you.

Okay. We're going to move to Measure 1660, same developers, ESRD patients receiving dialysis, hemoglobin level less than 9.

MS. SINGER: Okay, thank you. We have Paul Palevsky on the line, I believe, who's going to introduce this measure.

CO-CHAIR CROOKS: Paul, are you there?

DR. PALEVSKY: I was trying to get an open line, sorry.

CO-CHAIR CROOKS: Okay. We hear you. Go

ahead.

DR. PALEVSKY: Okay. I guess I now have the open line, so thank you.

This is Measure 1660, ESRD patients receiving dialysis, hemoglobin level less than 9 grams per deciliter. The description of this measure is percentage of calendar months within a 12-month period during which patients aged 18 years and older with a diagnosis of end stage renal disease who are receiving hemodialysis or peritoneal dialysis have a hemoglobin level less than 9 grams per deciliter.

Our rationale for this measure is that it is intended to look at the outliers at the low end of the hemoglobin distribution curve to identify patients --- to identify quality of care of patients for treatment of anemia.

This --- establishing a threshold for a measure such as this has been somewhat problematic. I appreciate that you just accepted the pediatric measure with the hemoglobin level at 10. This measure in a prior iteration had not

been endorsed by NQF using a hemoglobin level of 10, and we have revised this with a threshold at 9.

We are not contending that 9 is some absolute level at which a hemoglobin less than 9 is really bad, and above 9 is really good. That's not what we're trying to establish here. We're trying to have a way of looking at the less --- if you're looking at a distribution curve, at the left side tail with the contention that the larger the percentage of patients with a hemoglobin in that left-sided tail would represent lower quality of care.

What we have noted is that over the past several years there actually has been an increase in the number of patients with hemoglobins less than 9, and there are also data suggesting that there's an increase in transfusion rate associated with the changes in ESA therapy.

This, obviously, is an intermediate outcome measure. It is included in the RPA Kidney

Quality Improvement Registry for 2015.

The evidence supporting any specific threshold for --- at the lower end for hemoglobin, we recognize is problematic, but the KDIGO 2012 guidelines do suggest that for adult CKD 5D patients, ESA therapy be used to avoid having a hemoglobin concentration below 9 grams per deciliter by starting ESA therapy when the hemoglobin is between 9 and 10.

We know that from the Treat trial which was designed to give rescue therapy when the hemoglobin was less than 9, that that strategy of therapy of waiting until the hemoglobin of less than 9 was associated with an increase in transfusion rates. So, we think that this is an important safety measure for looking at the inadequate treatment of anemia. And, hopefully, you will find that this is a worthy measure where you did not find it so with a threshold of 10 when this was considered several years ago.

CO-CHAIR CROOKS: Okay. Thank you very

much. Our discussants are Michael Somers and Myra Kleinpeter. Who's going to go first?

DR. KLEINPETER: So, I'll start first. The summary provided gives a good idea of the rationale for the measure at this time. There's significant evidence coming forward that levels -- hemoglobin levels less than 9 can lead to worse outcomes, increased risk of transfusions, and the anemia guidelines from 2007 updates have a target range between 11 and 12, and then concurrently with the implementation from some of the data we have with the bundle payments there's been a trend downward of the overall average hemoglobin such that they're having higher proportions of patients with lower hemoglobins.

The first set of the information, also from other member comments, specifies the consequences of the lower hemoglobin may be indirect, and also increase in risk of transfusions which will impact candidacy for kidney transplantation.

Other information in terms of

performance gap, 38 percent of the patients do not receive the target hemoglobin supported by national performance, disparities exist in various populations, but there was insufficient data available for suitable comparisons. And at this time, some of the health disparities primarily exist among African American patients, and Hispanic patients. This may be multifactorial, but there are some disparities with care.

CO-CHAIR CROOKS: Okay. So, we're on the evidence discussion. Michael, did you want to add anything right now?

DR. SOMERS: No.

CO-CHAIR CROOKS: Okay. So, we open it to the Committee, discussing the evidence for Measure 1660.

MS. WAGER: I have a comment. When we talk about this less than 9, you mention adverse, you know, effects, events, or whatever. I don't hear one word about quality of life for the patient.

CO-CHAIR CROOKS: What would you like us to know?

MS. WAGER: Well, you don't have one, you know. Us professionals want the patient to continue working while on dialysis with a hemoglobin --- I remember dialyzing an 82 and my hemoglobin was around then. I couldn't work. I could just make a bed. So, I think that quality of life has to be an important component.

CO-CHAIR CROOKS: Lori.

MS. HARTWELL: I would just state, I would just double back. I remember I got two units of blood every six weeks for pretty much my whole entire age 12 to almost 23, and I would get a blood transfusion at a 26 hematocrit, and one time --- and I was telling --- I felt so bad at that time, I would drink Big Gulps before they drew my blood. And just to emphasize how awful you feel, and that's about an 8, 8.5 hematocrit, hemoglobin, excuse me. So, when you're at a 9, I mean, this is such a safety issue for patients, and I don't know if a lot of physicians today

understand the true effects of how bad you feel with anemia. And I think, you know, we get emails at our organization, patients tell us they don't feel good. They're not even aware their hemoglobin is making them feel bad, so there needs to be more awareness to this. And it is --- I mean, we're doing all this so patients can live the life they were meant to live, and hemoglobin and feeling good is such a big part of that.

CO-CHAIR CROOKS: Thank you. And I believe that --- well, that may be addressed in the body of evidence, is related to quality of life improvement. It should be. If it isn't ---

DR. KLEINPETER: Some quality of life information in terms in the references presented.

CO-CHAIR CROOKS: Okay. So, that is part of the evidence base, too, that we know that keeping it above a threshold, and this is a low threshold. Okay, Frank.

DR. MADDUX: So, I'd only make one comment. I mean, to me, this is a measure that's

a proxy for transfusion need. And the fact that there are no uniform transfusion policies in end stage renal disease patients makes it a little bit difficult to interpret how this might have unintended consequences.

My biggest concern about the less than 9 is the if used widely in the field, then suddenly it becomes a new target range. And that instead of saying that 9 to whatever, or 10 to whatever, it's pretty clear that the FDA guidance has been without a range. And in some way we might de facto create a range because of measures specifically being used, and that might be an unintended result of what is a reasonable concern here; and that is, that we begin to measure what happens for people that have very low hemoglobins. So, it's --- I'm quite conflicted on this particular area.

CO-CHAIR CROOKS: Thanks. Alan.

DR. KLIGER: Paul pointed out that there really is no consistent evidence for having a cutoff period --- a cutoff for hemoglobin for

adults. You guys did present some evidence for children. Having had the opportunity to review all of this for a publication last year, I can tell you that when you look objectively at it, looking at a whole population, that there really has been no compelling evidence that any specific cutoff, that any specific level, risk is increased below a particular level. Rather, it looks more likely that we need to be paying attention to individual patients.

You know, what we've heard from our patients here is absolutely right, that there are some individuals for whom a hemoglobin below 10 is impossible, and others who go and live really very comfortably with hemoglobins of 8 or 8.5. I have a real problem. And that's what the data show, so I have a real problem with a measure that goes across the board giving an arbitrary cutoff without any data suggesting that there are adverse consequences of the hemoglobin below that level.

CO-CHAIR CROOKS: Ishir.

DR. BHAN: Yes, I would agree with that. And, also, one of the comments earlier was that the goal was to --- well, one of the potential adverse effects of being below 9 in, for instance, Treat, was that there were more transfusions. I think if anything, an unintended consequence of this might be to increase transfusions, or maybe an intended consequence depending on how you look at it. But, regardless, an increase in transfusions as opposed to emphasizing EPO, there's no particular reason this would shift people more towards EPO or to an ESA than to transfusions. I think that's exactly that point.

CO-CHAIR CROOKS: Bobbi.

MS. WAGER: Dr. Maddux, thank you very much, because you spoke just how I was feeling when I read this measure. It scared me because I'm thinking oh, my God, now it's being lowered. There's going to be another target level. Thank you.

CO-CHAIR CROOKS: Lisa.

DR. LATTS: The other folks were first. Do you want --- are you going to --- okay. Well, I just have a question for the developer. If there's so much --- because this is not an area that I'm familiar with the science of. So, if there's so much science that shows there isn't a range, what's the developer's interest in creating a below number?

DR. PALEVSKY: So, while there is not --- we agree that ESA therapy should be individualized to the patient to maintain the hemoglobin in a range where the patients feel best. For the majority of patients, that's going to be a hemoglobin greater than 9. And I agree with --- I remember the patient in the pre-ESA era who were chronically transfused, iron overloaded, developed heart failure from their iron overload from the transfusions they received to maintain hemoglobins of greater than 8.

We are not --- and I tried making it clear, we are not implying that this number sets a range for treatment. This is set below the

level that most individuals would consider where you would treat patients with an ESA to maintain the hemoglobin level. I will say that since there has been removal of a target range, we have seen not only a downward trend in hemoglobin levels, but an increase in transfusion rate, which is the unintended consequence, I believe, of having eliminated a bottom level for hemoglobins.

CO-CHAIR CROOKS: Paul, I'm going to need to cut you off pretty soon.

DR. PALEVSKY: I understand the concern over this possibly being a stimulus to transfusion, but I suspect that it would not actually increase transfusion rates. Right now we're seeing the absence of a measure resulting in an increase in transfusion rates.

CO-CHAIR CROOKS: Okay, thank you. Stuart.

DR. GREENSTEIN: As a transplant surgeon, I would have to just chime in that my biggest concern when you put a number is that you will increase the transfusion rate, and it'll

make it much more difficult for me to ever transplant people. Especially, when I look at a patient, I will individualize what kind of hemoglobin they have in terms of when I can go a transplant on them. So, if I have a patient who's in his 30s and his hemoglobin is 8, you know, 8.5, I don't have a problem doing it, but if he's 55-60 years old, cardiac disease, I would have a big problem doing it. I mean, so I don't think you can give an absolute number. You have to individualize, otherwise you're going to sensitize everybody, and we're going to have a lot more people waiting, and waiting, and waiting for transplant.

CO-CHAIR CROOKS: Okay, thanks. Lori.

MS. HARTWELL: I just have a quick question, because I know in the previous years that I wasn't, there was a hemoglobin of 10 or below that was rejected. And I'm just having a little bit of difficulty understanding, because hemoglobins aren't being individualized to patients. Patients are way too low. I hear it

every day, and it --- you know, I mean, it's kind of unconflicted because I understand like you don't want to put a number on it. But then how do you make a safety issue so patients don't get too low? And they are too low in the community. So, I mean, I'm not an expert on it, but it would be nice if there was some kind of direction, maybe not from this body, the community that --- to not let people drop so low.

And just a comment to Alan, when I hear somebody being at an 8 or 8.5, I can only imagine that this patient is elderly and not moving around that much, because anybody who's younger can't really --- I can't --- I know many of my friends cannot move with a --- are just not very active at that low of hemoglobin. That's just my only comment.

CO-CHAIR CROOKS: Just to back up Alan, there is a lot of variability. There are patients who are quite comfortable, but many who are not. Lorien, I think you're next.

DR. DALRYMPLE: So, I was trying to

recall our discussion from a years ago, and I think the way at least I perceive this measure is it's a safety measure. There have been a lot of incentives put in place to insure hemoglobins don't go high, but we have not --- we no longer have lower boundaries for hemoglobin. And I think less than 10 is challenging in adult population because there's black box warnings, FDA, all of us who care for advanced CKD patients often have patients with hemoglobins of 9.2.

And if you believe the trials, that's probably acceptable, so I think, in part, my interpretation is the RPA responded to our last discussion of this measure saying okay, you're not comfortable with less than 10. Let's go to the guidelines and still try and develop a safety measure, which is really how I perceive this.

I don't know if it could incentivize transfusion. That's not how I perceive it. I perceive as perhaps adjusting EPO protocols and other things to insure we don't have a lot of these aids, and, hopefully, we individualize

therapy, and if people feel comfortable at a hemoglobin of 8, we leave them alone and don't give them excessive ESA.

But the RPA did submit supplemental material with the KDIGO guideline from 2012 that Paul mentioned with a grade of 2B. And I understand that's not a wonderful grade, but at least we do have a guideline grade now, which I think is more than we had a few years ago. So, from my perspective this is a safety measure to try and put a floor in, again. And it's not a target. I think it's like KT over Vs of 1.7. It's a floor, not a goal.

CO-CHAIR CROOKS: Okay. Back to Alan.

DR. KLIGER: Okay. Maybe the developers can help me with this. I see no evidence that as the hemoglobins have, indeed, come down as they have recently, I've seen no evidence of the community negative effects that you've described. Surely, there are individuals who are under treated, but in the absence of evidence of an ill effect by creating a single floor, rather than by

a showing that we're attending to each individual's needs, I see the risk as being greater than the potential benefit of setting up an arbitrary floor.

CO-CHAIR CROOKS: Okay. Any other ---

DR. PALEVSKY: Do you want me to respond to Alan's ---

CO-CHAIR CROOKS: If you can do it in 30 seconds or less.

DR. PALEVSKY: Alan, the only evidence that I can point to is the increasing rate of transfusion that has been reported since we have removed a floor.

CO-CHAIR CROOKS: Okay, thanks. So, I think the ---

DR. PALEVSKY: There's been almost a 50 percent increase in transfusion rate.

CO-CHAIR CROOKS: Correct. Okay. So, are we ready to vote on the evidence supporting this measure? Okay.

MS. OGUNGBEMI: The Committee is now voting for Measure 1660 evidence. Options are 1

high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are one vote high, nine votes moderate, six votes low, seven votes insufficient. Measure 1660 is in the gray zone for evidence.

CO-CHAIR CROOKS: Moving forward then, the gap.

DR. KLEINPETER: All right. So, for the performance gap, they have 38 percent of patients did not receive the target hemoglobin supported measure for a national performance goal. I already mentioned that there are some disparities regarding age and race, as well as gender, but there was insufficient data available for comparisons.

And in terms of disparities, some of those are that there are higher rates of anemia in African Americans. Why this exists may be multifactoral, but there still is a disparity in that number.

CO-CHAIR CROOKS: Okay. So, the gap is low, 5.4 percent, but we're considering this as a

safety issue. Who else would like to speak about the performance gap? Mahesh.

DR. KRISHNAN: Maybe just to give you some more contemporary data, I'm just looking at our data from March, and it's similar. It hasn't changed a lot since 2012 to now, so it's still about 5 percent, 6 percent. I mean, that's just the variance that there is.

CO-CHAIR CROOKS: Lisa.

DR. LATTS: And we do we know that 5 or 6 percent, based on the expert opinion of the people here, I mean, are those probably people that shouldn't be above 9, or are those people that probably should be there?

DR. KLEINPETER: So, at least in looking at our cohort and our population from both LSU and Tulane, a lot of those are the new patients that are presenting de novo to the ED who have had been in no care, and some of our sickle patients have sickle cell disease where our target overall is lower, so that we are keeping them a lower hemoglobin.

CO-CHAIR CROOKS: I believe the denominator exclusion includes --- I've looked at so many measures, but I think it excludes sickle cell anemia and other chronic anemia.

DR. PALEVSKY: It does.

DR. FISCHER: Just so I know what I'm voting on, so the performance gap that was characterized, does that exclude all the exceptions they have, sickle cell, cancer? What we just quoted, the 5 percent is not including any of those individuals?

CO-CHAIR CROOKS: I expect that's the case. Would you like to ask the developer? The quote, that's the measure. Alan. Put your card down, Alan. Okay. Josh first.

DR. ZARITSKY: So, when I read what they've provided in 1b.2, right, they say, you know, the gap here is 5.4 percent who had less than, you know, hemoglobin less than 9. I'm just --- you know, then they go on to say 36.5 percent of patients reported did not receive the optimal care. I guess that's referring to those not

within 11 to 12. Right? So, that's not --- I don't know why that's included in the gap.

CO-CHAIR CROOKS: Ishir.

DR. BHAN: So, just a question for Mahesh. You said you got a similar number, but you were not excluding people with chronic anemias and such, so that's all comers. So, presumably some of those people do have some other cause of anemia, and it's not really clear to me that this states that those are being excluded here, as well. So, I mean, ideally they should be, if that's what it's --- you know, if it's part of this document, but it's not clear to me that ---

DR. KRISHNAN: I can get into the specifications but I don't know how --- we can't exclude those patients either in our own data, because we don't have enough granular data to exclude them off those things, so I don't know you would do that, in practicality, for the other measure.

DR. BHAN: So, it may be as great as

5.4 or whatever the number is percent, but it may be lower than that, as well.

CO-CHAIR CROOKS: I think even including the exclusions, 5.4 percent is the noise of new patients coming in and patients, you know --- I conclude the gap is essentially zero, and it's kind of a quandary because you may believe it's a safety net, and it's an important safety net, but yet you have to vote on the gap. And the gap is not there, so it's a little bit of a conflict for me. So, with that in --- okay, Lisa, would you like to shed light on our dilemma?

DR. LATTS: No, no, not at all. It's just --- and Poonam and Sarah are gone, but it does almost beg the issue of some feedback to NQF, that for those measures that are safety measures it suggests a different criteria would be most appropriate than a gap for a performance measure.

CO-CHAIR CROOKS: So, let's vote on it, and then we can let them, you know, inform us if

we need to do something different. Mahesh.

DR. KRISHNAN: Just, you know, I'm looking now at the DOPPS National Sample Data, right, so the gap was as high as, in August of 2010, of 9 percent of 9s. Now it's about, sorry, 3 percent. Now it's as high as 5 or 6 percent, the gap has widened. Yes, widened over the implementation of the black box, which I think drove of this not bundled, but we won't go into that. But I think that that gap has widened, clearly, from 3 to 6 percent of the country.

CO-CHAIR CROOKS: You think the gap is increasing?

DR. KRISHNAN: Correct. In August of 2010, the sub-9 was 3 percent, and now it's 5 percent.

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

MS. OGUNGBEMI: The Committee is now voting on performance gap for Measure 1660. Options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are

zero votes high, five votes moderate, 16 votes low, and two votes insufficient. Measure 1660 fails on performance gap.

CO-CHAIR CROOKS: The pause that refreshes. So, we have voted that the performance gap is too low, even though this is really being considered as a safety measure. We're concerned that NQF maybe should have different criteria for evaluating a safety net measure, versus another kind of measure, or maybe not. Sarah or Poonam, can you ---

MS. BAL: In response to that, we do have a panel selected that will be meeting in the next couple of months. Helen mentioned it earlier, that will be working on if we need to change our criteria based on the type of measure, and possibly what the future use of it would be. But at the current time, the criteria stands, and it should be treated as if it was any other measure.

CO-CHAIR CROOKS: So, we can go on with our considerations. Although it's new, you know,

we --- it's in that borderline, maybe it's a reserve measure, maybe not.

MS. BAL: They would need --- have you guys discussed if you want to make it a reserve measure?

CO-CHAIR CROOKS: We haven't yet.

(Simultaneous speaking)

CO-CHAIR CROOKS: Yes, so we're in the gray zone, again, the double gray zone I'd call this.

MS. BAL: And before we move forward, Sarah and I disappeared for a little bit, mainly because we need to come up with a plan to get through all the measures. We are very far behind where we need to be. We really encourage you not to repeat anything what's already been said. Please just make any notes, you know, quickly. And, also, going through the measures, if you feel that reliability is good, no need to explain why.

Just say that you think the reliability is good, and we'll move forward. If

you ---obviously, if there are concerns, you should bring those up, but if the --- if you feel that the criteria is met and there's no further discussion needed, just state that and we'll move forward. You know, try to go as quickly as possible. We want to give the measures the due review, but we are very far behind at this point.

CO-CHAIR CROOKS: Yes, I'd point out that's not the Committee's fault that we're behind. I think we're doing a --- you know, I think --- I congratulate the Committee for doing its best job to fairly consider these measures. But with all that in mind, let's move on. So, we're now going to get --- so, do we move to the new measure then, or do we ---

MS. BAL: We go back to ---

CO-CHAIR CROOKS: Are we going to ask for --- we can't get reserve status if it's not ---

(Off microphone comment)

CO-CHAIR CROOKS: Okay. So, while that's all being considered, we're going to move

on to the next measure then. Why don't you take a measure, which one ---

CO-CHAIR ANDERSON: All right. The next measure is 1424, and the developers are Joe and Joe.

DR. MESSANA: My status has declined so much no one will sit with me any longer.

(Laughter.)

DR. MESSANA: So, in the spirit of brevity, I'm not going to make any specific opening comment except for to try to address a question that was raised in the work group. When I was reviewing the comments in preparation for this, someone asked -- one of the work group participants asked if the --- related to patients missing hemoglobin, this is pediatric patients, and whether these patients --- there was a correlation with other missing values from CROWNWeb, so we did a small additional analysis in CROWNWeb data.

Looking at the concordance between missing for calcium, phosphorous, and hemoglobin

from the CROWNWeb data, through each of the months of 2013, and there's a little bit of variation month to month, but generally there was concordance in that all three of those values, two bone and mineral intermediate outcomes, calcium and phosphorous, and hemoglobin were missing from somewhere between 35, and say 45 percent of patients in any given month from CROWNWeb in 2013.

There was concordance in that all three were present in similarly about 40 percent give or take from month to month. And then the remainder, which was somewhere between 10 and 20 percent in any month, there was a mixed pattern, a hemoglobin was there, calcium was not, hemoglobin was there, phosphorous was not, calcium and phosphorous were there, but hemoglobin was not. So, hopefully, that answers the work group question, and I'll stop there.

CO-CHAIR ANDERSON: All right. Bobbi and Myra, who's going first?

MS. WAGER: I am. Evidence shows that

it's a process measure and it's supported by KDOQI, how do you say it, I'm sorry, guidelines of 2006. There was a systematic review, and they rated it 2b.

CO-CHAIR ANDERSON: Myra, do you have anything else?

DR. KLEINPETER: Nothing else to add.

CO-CHAIR ANDERSON: Committee discussion? Josh.

DR. ZARITSKY: I'm going to include some other --- you know, in 1a.8(2), there's some additional evidence presented also. There's some additional evidence presented also in 1.8(2) that I think we've brought up before in a previous measure.

CO-CHAIR ANDERSON: Any further discussion? Let's call for the vote on evidence.

MS. OGUNGBEMI: The Committee is now voting on evidence for Measure 1424. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. The voting is open. The results are three votes high, 19 votes moderate, one vote

low, and zero votes insufficient. Measure 1224 passes on evidence.

CO-CHAIR ANDERSON: All right. Moving on to the performance gap.

MS. WAGER: On performance gap, based on 2013 CROWNWeb clinical data from January 2013 to December 2013, there's a quartile range of 22 percent, the mean and medium performance scores were 75 and 85 percent, with an inter-quartile range of 22 percent. Again, there's a gap due to the sample size used for the performance scores were considered too small to display useful disparities.

CO-CHAIR ANDERSON: Myra.

DR. KLEINPETER: Nothing else to add.

CO-CHAIR ANDERSON: Discussion on the part of the Committee? Call for a vote on performance gap.

MS. OGUNGBEMI: The Committee is now voting on performance gap, Measure 1424. Options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are two

votes high, 18 votes moderate, one vote low, and zero votes insufficient. Measure 1424 passes on performance gap.

CO-CHAIR ANDERSON: All right. Reliability testing.

MS. WAGER: Reliability testing demonstrates that the measured data elements are repeatable, on the measures collectively of CROWNWeb and the specifications described data element identification criteria, clearly. The measure is not risk-adjusted.

CO-CHAIR ANDERSON: Myra.

DR. KLEINPETER: Nothing else to add.

CO-CHAIR ANDERSON: Further discussion on the part of the Committee? Lorien.

DR. DALRYMPLE: Can I just ask one quick question. So, if children are hospitalized they're not excluded. Is that correct? And the only reason I bring that up is the N is so small in the pediatric practices, so the unit would be perceived not to have measured a hemoglobin if, in fact, it was because the child was

hospitalized. And maybe those hemoglobins get put in and counted.

DR. KLEINPETER: So, it all depends on when the unit measured the lab value. If they were hospitalized and did not receive the minimum number of treatments for that month, it would not be included, otherwise, it is included.

DR. KRISHNAN: It's not put in, though. So, there was a point in time over 2013 where patients were --- we were instructed to auto discharge patients, so if patients were in the hospital for a prolonged period of time in CROWNWeb, we were instructed to auto discharge them so that they didn't come up with errors.

So, the probability that that happened is there, and then on top of that --- I won't tell you specifically that labs are consistently reentered, right, because there's now a monthly close for CROWNWeb, so we have to get the data back in a certain period of time, then put it in before it's closed. So, there's a lot of ambiguity there.

And this is just --- my concern is this metric --- I know we've talked about CROWNWeb data transmission issues. This metric will emphasize those data transmission issues. Maybe that's what it is, it's measuring --- it's assuming that the data reflects reality and that the lab wasn't drawn, not that the lab was drawn and not transmitted, if that makes sense.

CO-CHAIR ANDERSON: Michael.

DR. SOMERS: I think at least for the children who are dialyzed in pediatric dialysis facilities, since the vast majority of them are associated with the center or actually physically in the center where they'd be hospitalized, I think you wouldn't lose those hemoglobin values. I can't comment on facilities outside of that.

DR. MESSANA: So, this CROWNWeb-based data, it's based on patient being admitted to the facility for a month, and Mahesh's statement regarding discharge for prolonged hospitalizations, if a patient was hospitalized, I think, for over a month, or some value like

that, they were discharged from the facility. Right? But short of that, they would be considered in the facility if they were hospitalized for 25 days. There might be a missing related to prolonged hospitalization and missed monthly blood draws.

Regarding the data transmission issue that Dr. Krishnan brings up, so whatever effect that has, dependent upon --- differing from organization to organization, when you compare adult CROWNWeb months with a hemoglobin value to the pediatric, there is a significant difference. The adults are much higher, so there's something well beyond this data transmission issue that Mahesh brings up. The gap is larger for the pediatric patients than it is for the adult patients.

DR. DALRYMPLE: But couldn't that be because in pediatrics just one patient will make your number look quite different. If you have an N of 12 in your unit, one or two children can drastically change your statistics. And adults,

you know, we have 100 patients, so ---

DR. MESSANA: I'm talking patient-level analysis, the total ---

DR. DALRYMPLE: The number of hemoglobins available, though. Is that correct?

DR. MESSANA: Yes.

DR. DALRYMPLE: So, a facility's percentage could quickly drop with just one child missing.

DR. MESSANA: It could --

DR. DALRYMPLE: Is that correct?

DR. MESSANA: -- but the statement that I'm making about a gap is a patient level analysis, national patient level analysis. There is something greater --- there's a greater shortfall of patient months with a hemoglobin in the pediatric population, not in the population of pediatric facilities.

CO-CHAIR CROOKS: Just one comment. This, unlike other pediatric measures, this does not exclude facilities with less than 11 patients. Right? This is all pediatric patients

which is, I guess, in its favor compared to some of the other ones.

DR. KLEINPETER: It still states that includes those that had 11 ---

CO-CHAIR CROOKS: 11 or more?

DR. KLEINPETER: 11 or more.

CO-CHAIR CROOKS: Oh.

DR. KLEINPETER: It's only 59 of the facilities had 11 people to report.

CO-CHAIR CROOKS: So, why would you need that, for just knowing that it was measured?

DR. MESSANA: Consistency sake in terms of small cells. We performed the analyses that way to be --- so that the information was consistent for you to consider.

CO-CHAIR ANDERSON: All right. Are we ready to vote for specifications and reliability?

MS. OGUNGBEMI: The Committee is now voting on reliability for Measure 1424. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are zero votes high, 22 votes moderate, one vote low,

and zero votes insufficient. Measure 1424 passes on reliability.

MS. WAGER: On feasibility, our data source for measure is CROWNWeb and ---

MS. BAL: Hold on.

MS. WAGER: I'm sorry.

MS. BAL: We're on validity.

MS. WAGER: I apologize. Again, data from CROWNWeb from January 2013 to December 2013, calculate facility level monthly and annual performance scores. Again, as Myra stated, it was 59 facilities and at least 11 eligible patients. And that included 1,280 patients in total during the testing.

Spearman correlation to assess the association between the annual performance scores and the NQF endorses standard mortality ratio using the 2013 SMR. Measure is maintained on the basis of face validity, the measure of hemoglobin as a dialysis quality measure was initially developed and approved by clinical TEP, which agreed that this is a quality measure.

CO-CHAIR ANDERSON: Myra.

DR. KLEINPETER: They also reported that the Spearman correlation coefficient was -.20 to P value .13, so those patients that had a higher percent of pediatric patients, the measures associated with a lower risk of mortality relative to those facilities with a lower percentage of pediatric patients where there's hemoglobin values measured. But it was not statistically significant.

CO-CHAIR ANDERSON: Any further discussion? All right. We're ready to vote on validity.

MS. OGUNGBEMI: The Committee is now voting on validity for Measure 1424. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are zero votes high, 22 votes moderate, one vote low, and zero votes insufficient. Measure 1424 passes on validity.

CO-CHAIR ANDERSON: All right, Bobbi, now you get feasibility.

MS. WAGER: Again, our data source for measure is CROWNWeb, the data is collected and generated and use by health care professional during the provision of care.

DR. KLEINPETER: Nothing else to add.

CO-CHAIR ANDERSON: Any further discussion by the Committee? All right, ready to vote on feasibility.

MS. OGUNGBEMI: The Committee is now voting on feasibility for Measure 1424. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are 14 votes high, seven votes moderate, one vote low, and zero votes insufficient. Measure 1424 passes on feasibility.

CO-CHAIR ANDERSON: Moving on, use and usability.

MS. WAGER: On usability, the information we got, it stated the measure is not currently in use, but is available for public use. On our conference call, I believe that someone had stated that Network 13 is collecting

data.

DR. KLEINPETER: So, nothing else other than we look at it among our pediatric population because it was a special project in our network.

CO-CHAIR ANDERSON: Any further discussion?

CO-CHAIR CROOKS: Are there --- I'd like to ask the developers, are there plans for using this in the future?

DR. MESSANA: I'm not aware of plans, but I'm also not aware of a lot of things that go on 7500 Security Boulevard, so ---

(Laughter.)

CO-CHAIR CROOKS: Well, let me ask the Committee Members, in the submission did they make a case for future use, public reporting or accountability? I don't think that they did. I have it here.

CO-CHAIR ANDERSON: John.

DR. WAGNER: I was just curious, since we have this high percentage of patients who are not entering --- who are not being tested for

either calcium, phosphorous, and hemoglobin, are there -- will there be unintended consequences? Is the reason that those tests are not being done that it is difficult for the patients to present for blood draws and-or difficulty in venipuncture? Do we understand anything about that?

CO-CHAIR ANDERSON: Any comments?

DR. KLEINPETER: So, just the only comment, that they're on hemodialysis, we get it as part of their regular monthly blood draw. The issue is that they're in the hospital, or if they're on home hemo and the parent or care giver forgets to collect it. Those are our missing values for our --- among our pediatric patients.

CO-CHAIR CROOKS: I just have to say that I'm puzzled by why you would want to go through the trouble of having a measure endorsed, and then there's no planned use for the measure. I'm a little puzzled what the reason for that is.

DR. MESSANA: So, I think the question is well taken. There have been a couple of things

that we've been assessing at CMS. The first is, you know, I think you've seen some of the consequences of just how to address assessing the pediatric population with measures, the minimum case size presents some issues any time you have a measure that is specific to the pediatric population, so there's that, and considering how best to use the measure.

I think the other concern is --- or not concern but issue is that there's an increasing focus on measures that assess performance of care as opposed to reporting of data. I think the distinction here, which causes some conversations internally to CMS, that are still kind of being hashed out.

In terms of the measure itself, my position on that would be that, you know, where there is a --- no clear gap --- where there is a gap in terms of data being available for a process measure like this, or for data being collected, I think it makes sense to have the measure available. It has been used for --- by

different --- by the networks and is available for use by other bodies, as well, regardless of its implementation by CMS. And I think that its availability for that use certainly continues to make sense.

CO-CHAIR ANDERSON: Mahesh.

CO-CHAIR CROOKS: You've got carditis, too.

CO-CHAIR ANDERSON: All right. I think we're ready to call for the vote on usability and use.

MS. OGUNGBEMI: The Committee is now voting on usability and use for Measure 1424. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are two votes high, 17 votes moderate, one vote low, and three votes insufficient. Measure 1424 passes on usability and use.

CO-CHAIR ANDERSON: All right. Is there any further discussion before we vote on recommendations for endorsement and the overall suitability for recommendations for endorsement?

All right, let's call for the vote.

MS. OGUNGBEMI: The Committee is now voting on overall suitability for endorsement for Measure 1424. Options are 1 yes/2 no. Voting is open. The results are 22 votes yes, one vote no. the measure passes meeting NQF criterias for endorsement.

CO-CHAIR ANDERSON: Moving on ---

CO-CHAIR CROOKS: Keep going. Obviously, you're better than I am. You're moving things --- obviously, you're the better facilitator for this, so please go ahead with 16 ---

CO-CHAIR ANDERSON: No, 2699, and it's Joe. Are you by yourself again?

DR. MESSANA: No. Dr. Doug Schaubel flew in today. He's the biostatistician who is the lead statistician on the transfusion metric, so he joined me in case there are any technical questions. I would like to take the two to three minutes allotted to developers for this measure, because I think I have a bit of perspective and

history that may help answer some of the questions that were raised during the work group. And I prepared a written statement.

The transfusion ratio measure is a direct product of the 2012 anemia TEP that I facilitated. TEP members included Jeff Berns, who is the Chair, Klemen Meyers, John Stivelman, the late Kathe LeBeau, Diana Hlebovy, I always mispronounce her name, I apologize again to her. Sheila Doss-McQuitty, Dr. Robert Kane, and Dr. Harvey Luxemburg both from the Food and Drug Administration. They were non-voting participants of the TEP. There was a pediatrician who was supposed to participate but couldn't because of a family emergency.

So, the context for the 2012 TEP, as most of you know, included emerging safety concerns about the use of aggressive hemoglobin targets and high doses of ESA at approximately the same time that a fundamental revision to the Medicare ESRD PPS bundle was made that shifted financial incentives and clinical incentives from

more use of ESA to less use of ESA. The TEP's concern was largely in protecting patients from both overuse and under-use of ESAs, and I think some of that discussion from the RPA measure discussed a few minutes ago kind of reflects some of those concerns.

Also, the 2011 revised FDA package insert for EPO, which included a statement about transfusion avoidance being one of the major goals of EPO therapy in this population was reviewed by the TEP. The deliberations. First, the TEP reaffirmed the dialysis facility central role in anemia management. They recommended three measures, hemoglobin less than 10, hemoglobin greater than 12, and a risk-adjusted transfusion measure.

Regarding the relationship between anemia management and transfusion, the TEP was impressed by the impact the introduction of ESAs and the evolution of anemia management practices have had on transfusion rates in dialysis patients from the initial reports of Eschbach and

Neil Powe over two decades ago, the secondary outcome from some of the larger clinical trials, the Canadian and European trial, in particular.

They used the paradigm that the decision to transfuse a patient is based on clinical judgment, but that that clinical judgment depended on both the patient's achieved hemoglobin and the clinical context, be it planned surgery, unplanned surgery, GI bleeding, et cetera. They felt that the clinical context was often not under the dialysis facility's control, but the hemoglobin concentration often was.

A major area of discussion related to which comorbidities and events should comprise the exclusion list. The TEP recommendations for exclusions were followed in development of this measure. They settled on an exclusion list that reflected conditions that were very likely to result in decreased ESA responsiveness, and were also conditions that may well reflect either strong relative or absolute contraindications to

ESA use in many cases.

They explicitly chose not to recommend inclusion of other diagnoses, such as GI bleeding, and other conditions to the list. Regarding the identification of transfusion events in Medicare claims which has come up in a work group discussion and public comments, the TEP should not be held responsible.

That was an implementation issue, and we developed the methodology, based on American Red Cross transfusion billing guidelines published within the last several years. This technique has been used extensively by Dr. Collins and colleagues to publish several research papers in peer reviewed clinical journals over the last five years, and to support Dr. Collins' presentations at national meetings on the topic of increasing transfusion rates subsequent to 2011.

The methodology's validation has been limited to one article that I could find from 2000 or 2001 from Dr. Powe's group describing the

high specificity and moderate sensitivity of identifying transfusions through these claims when compared to electronic medical record system review in a single large referral medical center. I'll stop there.

CO-CHAIR ANDERSON: Thanks. And the discussants are Dodie and John. John.

DR. WAGNER: Thank you. I will lead off. So, we have the standardized transfusion ratio, and the numerator statement is the number of eligible observed red blood cell transfusion events, and they are defined, as we've just heard from revenue codes, procedures codes. And the denominator is the number of eligible red blood cell transfusion events as defined in the numerator statement that would be expected, so there is a statistical modeling of the expected events, and a ratio is created around those.

The evidence to support this is based on KDIGO in 2012. There are three guidelines cited, two of them have to do with reducing blood transfusions as a goal with 1b evidence, and one

of them has to do with managing chronic anemia without excessive risk of ESAs, and, therefore, in such instances where there's a concern, red cell transfusions may be preferred. And that evidence is rated 2c.

There were also 17 articles reviewed by the TEP, and many of those have to do with issues related to the reduction in transplant suitability resulting from aloe sensitization, and hemoglobin target analyses. So, there really is not the risk tradeoff as to what would be the ideal use of ESA versus the risk-benefit ratio of transfusing is not cited in any of the literature. So, that is the evidence. As far as the performance gap goes, there is a listing ---

CO-CHAIR ANDERSON: John, I think we'll just ---

CO-CHAIR CROOKS: Yes, hold on a minute.

CO-CHAIR ANDERSON: Yes, we'll just do the evidence for right now.

DR. WAGNER: Okay.

CO-CHAIR ANDERSON: Dodie.

DR. STEIN: There was one statement about there are no empiric data per se, but transfusions are considered suboptimal care, therefore, the adverse outcomes and things.

CO-CHAIR CROOKS: Okay. So, this measure is claimed to be an outcome measure and, in fact, this is an outcome measure. I think it's the only one that's been presented to us that really is an outcome measure in the sense that it's --- they didn't claim it to be an intermediate outcome.

So, the burden of evidence is different in an outcome measure, and it's not to show --- link it to mortality or other outcomes, but it's to show that it can be changed by something up stream, a process, an intervention. And I don't think they approached it that way, so it left me a little confused. So, I'd like to ask developers what you were thinking about, or am I misreading your application?

DR. MESSANA: Well, maybe we weren't as

explicit as we should have been. There were --- the TEP was presented with information about achieved hemoglobin and subsequent transfusion risk in the subsequent quarter after an average hemoglobin in the preceding quarter, which showed an inverse relationship between transfusion events and that prior achieved hemoglobin.

In addition, since the TEP, and as part of the submission, Dr. Collins presented a facility-level publication in 2014 in American Journal of Kidney Disease dividing achieved facility-level hemoglobin by quintiles and looking at transfusion risk, showing that that achieved hemoglobin was predictive of subsequent transfusion rates at the facility. That's the only additional evidence that I would offer to suggest that there is an intermediate outcome that is fairly significantly associated with transfusion events.

CO-CHAIR CROOKS: So, maybe making the case that how you manage anemia will impact this outcome?

DR. MESSANA: That was the belief and the intent of the TEP when they recommended this measure.

CO-CHAIR CROOKS: Just like mortality as an outcome, we know zero mortality isn't achievable. Is there a target transfusion ratio that we --- the units should be striving for? Is one the right ratio?

DR. MESSANA: Well, one is --- describes the average, if you will, use of transfusions by peer facilities in a risk-adjusted rate within that particular year, so it's like the standardized mortality ratio, standardized hospitalization ratio, it is a relative ranking of transfusion use in that year, adjusted for patient and facility characteristic.

CO-CHAIR CROOKS: Right. So, you know, I'm --- I know that mortality, it's better to have less, you know. And is there a --- is it best to have no transfusions, or have a zero? I mean, what is the target, if this is an outcome? Which way are we trying to move this outcome?

DR. MESSANA: I think that that's a question that's more appropriate for the group, but since you directed it at me, I would suggest that lower rates would, generally, be better. There may be some exceptions, particularly in patients that are not good candidates for ESA use, and I think for the most part those are addressed by the exclusions. About 20 percent of patients are covered by exclusions in any given year, 19-20 percent, something like that.

CO-CHAIR CROOKS: To help you do your work, I do think this is an important outcome. It's an important measure, and I think the case that you should have probably made to us is this outcome, standardized transfusion ratio is related to how you manage anemia, processes up stream do affect it, and to make that link for us in a very clear way. Then I think the discussion on evidence is over. It's not --- it doesn't have to link to other outcomes. So, I'll stop beating that horse, and open it up to the floor for discussion of the evidence.

CO-CHAIR ANDERSON: Alan.

DR. KLIGER: Just very quickly, it's hard for me to imagine this as an outcome measure. It's either an intermediate outcome, or a process measure. We're not seeking to have a ratio. I mean, that's not an outcome.

Furthermore, the standardized measures, Peter, as you pointed out, we have to be very careful about understanding what those mean. A standardized hospitalization or mortality ratio, the direction of that, and the goal of that looks fairly clear. But if you talk about things like a standardized infection ratio, for example, we'll be talking about that later, achieving the same level that everybody else in the country achieves, gosh, I don't want to be judged if my medical center, that we're so good with infections that we're just as bad as everybody else.

The standardized infection ratio doesn't make a whole lot of sense to me. A standardized transfusion ratio is only useful, as

you point out, to understand what direction you're going in, and why. I'm not certain that zero transfusions are right. I'm also not certain that high transfusions are wrong. I don't know. So, I think this is sort of a process measure, and I think we should judge it according to the standard route for a process measure.

CO-CHAIR ANDERSON: Josh.

CO-CHAIR CROOKS: Well, they did claim ---

CO-CHAIR ANDERSON: Oh.

CO-CHAIR CROOKS: I'm sorry, just to finish the conversation.

CO-CHAIR ANDERSON: Sorry.

CO-CHAIR CROOKS: They do claim it's an outcome measure, and even in the evidence section they checked the box outcome measure. And I can see it that way, I can see it as parallel to standard hospitalization ratio, mortality ratios, and so on, but that's the case they're trying to make, I think.

DR. KRISHNAN: I have a question on

that very specific aspect. When we started this conversation, we said that two of the measures, SHR and SMR were moved into the future because they're looking at the --- sorry, because they were moving the SMR and SHR further up, because we're looking at the risk adjustment methodology. Is this the same risk adjustment methodology, or different?

DR. MESSANA: No. There --- so, the risk adjustment strategy for this metric, in addition to incident comorbidity is 27-28 comorbidities which are included in the first level, the patient-level metric. The risk adjustment strategy is the exclusion of approximately 20 percent of patients with comorbidities from prevalent comorbidity claims that are associated with malignancies, sickle cell anemia, other acquired --- excuse me, other hereditary hemoglobinopathies, so in addition to the incident comorbidity adjustments, there are prevalent comorbidity adjustments from Medicare claims.

DR. KRISHNAN: So, this is sort of a hybrid between the two?

DR. MESSANA: Well, it has additional ---

DR. KRISHNAN: So, will the deliberations that you have for the SMR and SHR affect this risk adjustment methodology, as well, or should we not think that way?

DR. MESSANA: Don't know. The SHR and SMR TEP is charged with addressing those. I suppose that depending upon where that discussion goes, who knows. I can't predict the future, and I don't think you should make any decisions based on that statement.

CO-CHAIR ANDERSON: Okay. Josh.

DR. ZARITSKY: I have a real problem with this as an outcome measure, and that really affects how you're going to judge the evidence. And going toward the evidence, you know, for the denominator, you know, what's the evidence that that denominator is valid in this --- I mean, there should be some evidence presented that

there's --- that that denominator has some validity.

I guess I'm asking for evidence that there is --- you know, what is the evidence that this is a valid measure, that there is a standardized expected transfusion ratio in this population? I mean, I could stop with number one, the outcome. You know, I want to know is it an outcome versus intermediate, and then I'm just having a hard time --- I'd like to see some evidence here using this standardized --- maybe I just don't know it because I'm not an adult nephrologist, that this standardized transfusion --- expected transfusion has validity --- you know, there's evidence for its use in this population.

DR. MESSANA: So, Doug, do you want to try to comment? I'm not sure I understand the question, but Doug apparently does.

DR. SCHAUBEL: I'm not totally sure I understand either, but just a comment pertaining to your question, is that the --- so, every

center is observed, transfusions are compared to the expected, and that's the same general procedure as what's used in mortality and hospitalization. And, ideally, perhaps you'd like to be much lower than expected, and perhaps what happens at the national level is not really the target because it should be lower in the nation as a whole. But at the end of the day, I'd say as a target, it's a way to compare each center in a uniform way.

DR. ZARITSKY: How do you --- just tell me since I don't know, how do you calculate the expected, then?

DR. SCHAUBEL: Yes. So, it's considering the patient characteristics and the national --- there's a baseline rate which is independent of patient characteristics which applies to everybody, which applies to every patient across the country. And then that patient's characteristics scale up what the expected is for that patient, either scales it up or scales it down, depending on what those

characteristics are. And then the sum is taken for every patient within that center to come up with that expected count, which is then compared to the observed. That's the same --- in a general sense, that's the same as the SMR and the SHR.

CO-CHAIR ANDERSON: Frank.

DR. MADDUX: So, I'll ask Joshua's question a little bit differently, but similarly to colleagues on the Committee and others. I don't understand why a standardized ratio is appropriate for this particular measure, because I believe that if we're looking at trying to understand both the distribution of transfusions and the rationale for transfusions, I don't understand how this particular measure actually gives us complete clarity, other than we know that you can rank them across a distribution against an expected range.

And I can tell you without any transfusion guidelines, unlike there may be in hospitals for cardiac surgery, I know of no transfusion guidelines for end stage renal

disease patients. It's very much a one by one decision, so I'm just struggling with the rationale of creating the standardized ratio and the evidence that supports in another way what an expectation should be.

CO-CHAIR ANDERSON: Andy.

DR. NARVA: Yes. I think, you know, what this reflects is an assessment of management of anemia, and we're often looking for ways of improving quality of care while at the same time promoting individualized decisions. And I think that this sort of measure, which really gets at avoidance of anemia, avoidance of transfusion really has that built in. And I think the lack of a guideline on transfusion isn't necessarily a negative thing. Here I think that this allows --- some patients clearly are candidates for transfusion, some patients aren't.

In general, we try to avoid transfusion and that's --- even the FDA agrees with that, so I think this sort of is --- embodies the opportunity to improve care while

at the same time promoting sort of individualized patient decisions. And I don't know how that fits into the outcome versus process measure, but I think it allows us --- in some ways it may be the best reflection of quality management of anemia.

And I also think --- you know, I'm sure many people here have been on medical review boards and networks, but these ratios are extremely good at identifying outliers, at least for the examination of quality issues. And I think that's an important purpose, presumably, of this measure.

CO-CHAIR ANDERSON: Lorien.

DR. DALRYMPLE: And I think I have more of a procedural clarification question, for perhaps Sarah or the Chairs. So, have we come to agreement on what type of measure this is, as to whether it's an outcome or an intermediate clinical outcome? Because I think that would then help the Committee have an active discussion about whether we're grading, or voting yes/no.

And then when there are these

methodologic issues, especially with standardized ratios which are inherently complicated and risk-adjusted, should we be conferring some of that discussion to the reliability and specifications where we all argue about the merits of biostatistical techniques to report meaningful data, of which there are many choices, or do we need to be discussing that in the evidence?

MS. SAMPSEL: Okay, so that's two questions. First question, process or outcome? The developers have submitted this as an outcome measure. They did attempt to provide a link, which is the NQF requirement between a process of care and an outcome. They are indicating the outcome that they are, you know, assessing measurement of as transfusion, so, you know, I think that Dr. Narva said it correctly, that this is kind of management of anemia to prevent transfusion as the outcome. Correct me if I'm interpreting that incorrectly.

DR. MESSANA: I think that reflects our opening statement, and the introductory statement

in the application.

MS. SAMPSEL: So, you are --- you will when you vote on this, you'll evaluate this based on the question, does the rationale that was provided support the relationship of the health outcome of a transfusion to at least one health care structure, process, intervention, or service as provided in the evidence?

Your second question regarding two discussions on the methodology, the statistics, that fun conversation I'm sure is going to happen under reliability and validity.

CO-CHAIR ANDERSON: Michael.

DR. FISCHER: Just going back to what Frank said, and if you look at, again, the algorithm we used, you know, if it is an outcome measure, then really it's pass or no pass. And I think there's no ambiguity when you think of SHR and SMR, in terms of the opinion that higher is worse.

But this is a case where I think that there's a lot more ambiguity about higher --- if

you're higher, if your standardized transfusion ratio happens to be higher, is that unambiguous that that reflects a problem with quality of car,, and that there's a process or a structure that needs to be changed to improve quality? And I guess that's why I have --- I distinguish this from the aforementioned two ones that I mentioned because, to me, I think there's a lot more ambiguity and uncertainty about how it's interpreted, and, hence, what a facility is supposed to do in terms of quality of care.

CO-CHAIR ANDERSON: Lisa.

DR. LATTS: Well, and it's sort of --- Michael, in response to that, I think that it is one of those measures that you have to get away -- you can't think of it as higher or lower, as better or worse. It's only an observed versus expected, and it's very --- that's why it's critical that whoever is doing the measuring then does the expected accurately, and you as a facility, or whoever is submitting it, won't know until you get the data crunched whether --- how

you performed.

Because you won't know until you know how your relationship is from observed to expected, because it's not supposed to be low, it's not supposed to be high. It's your relationship of that observed to expected that is the key outcome. And it's a very uncomfortable place to be, because you don't know if you're supposed to be high or low.

CO-CHAIR CROOKS: I'd like to suggest a way forward, and that is let's accept the application that this is an outcome measure, and let's vote based on the evidence, based on that they have or you believe there's a linkage to a service, an intervention, health care structure, or process. And let's just vote on it that way, because that's what they're asking us to do. It's not necessary for them to link it to other outcomes. And in my opinion, the case is made, or I'm going to give them that in my own personal opinion. But I think we should just vote on it that way, and then we can move on from there.

Does that make sense?

CO-CHAIR ANDERSON: I'd just like to make one other comment, and that's where it's at a facility level, this measure is at a facility level. And dovetailing on what you said, Michael, I'm not sure that's where the right place is. If this shouldn't be at a clinician level versus the dialysis facilities, because many of the facilities aren't doing standardized, or blood transfusions, within their facility. They're sending their patients to the blood bank, or wherever, so I have concerns about it being a facility-level measure.

DR. MESSANA: The reason that this is a facility-level measure, and the TEP felt very strongly about this, dialysis facilities are held responsible under the CFC 494 regulations for anemia management. Dialysis facilities are paid for anemia management, and are the sole source for administration of ESAs in chronic dialysis patients on Medicare. And the TEP felt very strongly that anemia management contributed to

the decision to transfuse because achieved hemoglobin in, you know, different clinical context was a main contributor to a decision to transfuse given the situation, or given the clinical context. That's why they supported this or they recommended this as a dialysis facility level metric.

CO-CHAIR ANDERSON: But I guess I would also argue that it's the clinician's decision to transfuse. It's not the facility's level to make that decision whether or not to transfuse the patients.

DR. MESSANA: I accept that, but the clinician uses --- it's the accepted standard of practice for me to transfuse a patient who has some absolute achieved hemoglobin threshold, and the achieved hemoglobin threshold is at least partially in most situations under the facility's control and responsibility.

There are situations where the achieved hemoglobin that you would choose to --- that I, as a clinician, would choose to

transfuse is going to be different than an absolute that Red Cross would set based on active GI bleeding or other conditions. But still the achieved hemoglobin in almost every decision I've ever made to transfuse is a major contributor, a major clinical input, to that decision.

CO-CHAIR CROOKS: Lisa.

DR. LATTS: Just one quick sort of comment and, again, from sort of an outsider perspective as I'm looking at this, is that this might be a nice compromise if you can operationalize it, which maybe you can't. If you don't know what the hemoglobin targets are, if they should be 8, 9, 10, or something, or none if we can't come up with a target, maybe an observed to expected is a better way to measure it.

CO-CHAIR ANDERSON: All right. Are we ready to call for the vote based on evidence?

CO-CHAIR CROOKS: We're linking the outcome to process, structure, intervention.

MS. OGUNGBEMI: The Committee is now voting on evidence as an outcome measure. The

options are 1 yes, 2 no. This is for Measure 2699. Voting is open. The results are one or 12 votes yes, 10 votes no. The Measure 2699 is a gray zone measure for evidence.

CO-CHAIR ANDERSON: So, performance gap. John.

DR. WAGNER: Okay. So, gee, I was hoping --- oh, well. So, there is based on the STRRs reported from 2009 through 2012, a performance gap in comparing, for example, inter-quartile range of about 0.6 to 0.7 each year. And there were also analyses of disparities, there was some statistically significant changes in transfusion ratios obtained when viewing females, Native Americans, Asians, Blacks, other race versus --- and Hispanics, which were not felt to be clinically significant in that when one adjusted for those, the STRR seemed to have the same results.

CO-CHAIR ANDERSON: Dodie, anything else? Any other comments for the Committee?

CO-CHAIR CROOKS: The only other

consideration for performance gap is you sort of need a target to say there's a gap, and there's a dispersion of results. And if the idea is that's a gap because everybody should be coming together close to one, okay, but then that suggests that the target is one. So, gap is a little hard to interpret. I do agree that there is dispersion of results.

CO-CHAIR ANDERSON: Okay. Mahesh.

DR. KRISHNAN: I just --- one thing I didn't understand, maybe get the Committee's input. If the standardized, mean standardized transfusion rate hasn't really changed, if I read this correctly, all that much from 2009 to 2012, you guys just described the mean hemoglobin has gone down, how do we reconcile those two?

(Off microphone comment.)

DR. KRISHNAN: But we're saying it's a surrogate for hemoglobin, but we know that the mean hemoglobin has gone down. If the mean hemoglobin has gone down for the country and this measure hasn't changed, how does that --- that

doesn't seem to correlate.

CO-CHAIR ANDERSON: Alan.

DR. KLIGER: A transfusion ratio just describes variation around the mean of the country. And it's hard for me to understand what a performance gap means given that kind of a measure. I don't see how we can interpret it a performance gap.

CO-CHAIR ANDERSON: John.

DR. WAGNER: Yes. So, I agree that if you don't know what your target is, it's hard to understand how one should view these dispersion of values. And the fact that the issues that were thought to provoke a concern around this, namely, the change in FDA indications in prescribing information and the impact that the billing reimbursement methodology would have on transfusions doesn't seem to be reflected in any real change in the dispersion of these values. But, again, it depends on how one views the value of the STRR as a clinical outcome.

CO-CHAIR ANDERSON: Josh.

DR. ZARITSKY: If you just use the gap to say hey, well it's just --- there's this variation and stop there, and say well, that's your gap. There's a variation, because we don't know what's better or worse, but there's a variation. It's not like we're all at one, so there's no point in --- so, there's always going to be a gap.

(Off microphone comment)

CO-CHAIR ANDERSON: Turn your mic on, Alan.

DR. KLIGER: But whenever any of these standardized ratios -- by definition, will just describe the variation around the mean, always will be there, and I just don't know how to --- what a gap means in that context.

CO-CHAIR CROOKS: I would suggest as sort of a flaw in the way we're supposed to consider these, if it is an outcome measure, then the gap doesn't mean the same thing, and maybe gap shouldn't be part of the consideration for an outcome measure. But there we are, we have it

right now. Sarah, would you have anything to ---

MS. SAMPSEL: I don't have anything to add.

CO-CHAIR CROOKS: Okay.

(Off microphone comment)

CO-CHAIR ANDERSON: Very brief, because we're going to have to go to public comment.

DR. MESSANA: So, the only issue is that absolute transfusion rates increased after 2011. You're looking at risk-adjusted rates here, so in terms of justifying the need or the protection for the population, the absolute transfusion rates may be more important to consider.

CO-CHAIR ANDERSON: All right. With no further discussion, let's move to vote on performance gap.

MS. OGUNGBEMI: The Committee is now voting on performance gap for Measure 2699. The options are 1 high, 2 moderate, 3 low, and 4 insufficient. Voting is open. The results are one vote high, 10 votes moderate, three votes low,

and nine votes insufficient. Measure 2699 falls in the gray zone for performance gap.

CO-CHAIR ANDERSON: We're going to need to pause on this measure for public comment.

OPERATOR: If you'd like to make a comment please press *1. There are no comments at this time.

MR. DIAMOND: I would just urge the Committee to ask staff to provide the definitions for structural, process, and outcome measures. I was totally confused about the discussion that this last measure is defined as an outcome measure. My understanding, and I may be totally incorrect, of an outcome measure is a description of a patient's status, and the change in the patient's status.

Mortality is an outcome measure. This is a process measure. This is something that has been done to a patient. If this becomes an out -- - if this is defined as an outcome measure, then anything we do to a patient and any ratio that we attach to doing something to patients becomes an

outcome measure. So, I think the way to resolve this would just be go back to --- I'm sure there's some standard definitions within the archives of NQF that define something called a structural, process, and outcome measure.

CO-CHAIR ANDERSON: Did you introduce yourself for the record?

MR. DIAMOND: Oh, I'm sorry. For the reporter, thanks, Lou Diamond speaking for myself.

CO-CHAIR ANDERSON: All right. Let's continue on our measure --- turn your mic on.

DR. DALRYMPLE: Would it be possible to maybe take a short break. I think this is actually a fairly complicated reliability section to go through, even if it's just for five minutes for the Committee, or is that not a choice? I don't perceive this to be a straightforward reliability, in large part because of the statistical methodology, and all of the issues that have come up today. And it's ---

CO-CHAIR ANDERSON: Actually, what we

would like to propose is to go ahead and get up, and get lunch, but actually do a working lunch so we can get through more of the measures. So, if you could be back by 12:30, and then we'll do that. Thanks, Lorien.

(Whereupon, the above-entitled matter went off the record at 12:15 p.m. and resumed at 12:30 p.m.)

CO-CHAIR ANDERSON: All right. We're back on Measure 2699. And we're at specifications and reliability. Dodie or John?

DR. WAGNER: Okay, well pardon my chewing.

So, specifications, the numerator includes claims data and various procedure and revenue codes. I guess the comment I would make about that is that it seems there are certain procedure codes that will allow one to capture the number of units transfused versus revenue codes which I think only give a thumbs up or thumbs down on a transfusion event.

So, I'm not sure if that affects the reliability of the data. There's no information about that. And if I'm confused about that point, if someone can clarify that, I would appreciate it.

Okay, as far as the denominator exclusions, there are the exclusions that are thought to influence anemia management in a way that would make it difficult to use ESAs and, therefore, the things that one has seen in other metrics have been put forth as exclusion events.

However, I think unlike other anemia metrics, the only hemoglobinopathy that is mentioned is sickle cell and there is no other exclusion for other hemoglobinopathies and, depending on the coding, I'm not sure if that would then result in some confusion, for example, someone had SC disease or something along those lines.

In terms of other exclusions that would relate to the KDIGO Guidelines that were cited, ineffective bone marrow could be the

result of other issues such as inflammation, surgical trauma and those events are not exclusion events.

As far as the statistical modeling which I think will be a large discussion, this is using a statistical model which is risk-adjusted for age and for nursing home status at time of the ESRD onset and years on ESRD, diabetes status and comorbidities as at the time of filling out the 2728.

And the risk-adjustment does not get updated as one accumulates time on dialysis. And this risk-adjustment is then put into a Cox model to come into the expected transfusion rate ratio. I'm sorry, the expected transfusion rate and then observed over the expected delivers the ratio.

So, I think there is, again, a question as to whether one can view an expected rate as indicative of good quality care or consistent with recognized standards of care and whether there should be other risk-adjustments within the model that speak to issues such as

bleeding events, such as surgical events, such as infectious events and those have not been incorporated into the model.

And also locus of transfusion since there are some dialysis facilities which have the ability to transfuse and many which do not. It's not clear what the relationship of the locus of transfusion might bear on this model.

So, again, another way of looking at the reliability of the data are to discuss the inter-unit reliability and based on their analysis of the inter-unit reliability, the developers indicated they believe there was a moderate degree of reliability.

So, that is my comments. I think this is very problematic to claim that this is a reliable indicator of quality.

And, with that, I'll open up the floor.

CO-CHAIR ANDERSON: Dodie, any additional comments?

DR. STEIN: The only other comment is

I may mention the larger facilities are going to have a higher IUR.

But I just have problems with the whole calculation of all of this and then some feasibility issues for later.

CO-CHAIR ANDERSON: Right. Alan?

DR. KLIGER: I guess I'm not surprised that there should be a correlation between a standardized transfusion ratio and hospitalization or mortality. Maybe we're just identifying sicker patients. I'm just not sure what that tells us. I don't know that that validates the measure, it just identifies perhaps a group of sicker patients.

CO-CHAIR ANDERSON: Lorien?

DR. DALRYMPLE: And I think -- I'm not sure if we're going to discuss this as a committee or further pose it to the --

I'm not sure if we're going to further discuss this as a committee or move it to the developers in the name of time, but this issue of how transfusions are counted, depending on if

you're a revenue center or a procedure code, it sounds like from this submission, and I do greatly appreciate the transparency on the approach to counting, the procedure codes tend to be a smaller percentage but I don't have an understanding of what percentage.

And then this question comes to my mind, well, Facility A is associated with Hospital A and Facility B is with Hospital B and Hospital A tends to use revenue centers, so there's going to be less events systematically counted compared to B using procedure codes.

Is Facility B really just being penalized because they're associated with hospitals that use different billing practices and does this happen systematically?

Because my understanding is all inpatient claims are counted unless you are a transplant hospitalization.

So, we're now making the facility accountable, not just for their care of practices, but the care of practices at the local

hospital may have very different transfusion standards than we have and may also have systematically different billing practices.

So, I don't know if we, as a group, should discuss all these issues or, in the name of time, ask for clarity? Peter? Connie?

DR. MESSANA: Well, so, Lorien, I'm not sure I can provide a lot of clarity on the latter question. It is certainly, as I pointed out in my introductory statement, this approach is the best one can do with the claims basis.

But there certainly is opportunity for a hospital to hospital variation in how they, you know, how they submit those and whether they submit those because inpatient hospitalizations are DRG related on whether they include a specific line item.

When we've looked previously at the association between dialysis facilities and hospitals, the paradigm that you develop, it's much more complicated than that.

Most freestanding dialysis facilities

have relationships with more than one hospital. There's a subset that there's a geographic singularity, but it's much more complicated than that.

And I would say I think the majority of facilities have relationships with more than one hospital, although geographic issues may still come into play.

The reason I had the card up is in the reviewers statement, I wanted to correct one, what I think, is factual omission. There are a number of hereditary anemias, hemolytic anemias, that are included in a bunch of sickle cell variants, sickle cell C diseases and others.

So, there's a whole family from the categorical -- the hierarchical coding system, CMS coding system of hereditary anemias, not just sickle cell that are part of the exclusion list and the code table that was included with the submission.

CO-CHAIR ANDERSON: So, just a quick follow-up. Have you empirically tested if you

treat procedure codes the way you treat revenue centers, if any findings for facilities change on their O/E?

DR. MESSANA: We have not. We have a breakdown of revenue center codes that are isolated that don't have a procedure code with them and how much overlap there is. And a sizable minority are based on revenue center codes only.

So, that's why we were careful to label transfusion events because we probably are systematically undercounting units of blood transfused with this methodology. We can say one or more blood transfusion was administered but on this day and only if we have granularity from procedure codes can we say that for sure that were two or more or two or multiple transfusions provided on a day.

CO-CHAIR ANDERSON: But for those patients, you have granularity available. I just want to make sure I understand the numerator. You will use it in those cases.

So, there's patients that don't have granularity so we just say, okay, you're one. But this patient where procedures were systematically, you could give them a six, you could give then an eight if they got eight units of blood because you -- although I think you can only do one procedure per day, correct?

So, it'd be a max of one unit per day. But if you're in the hospital for three weeks and that hospital uses that schema, you can get up to one a day. And then --

DR. MESSANA: Right, it's one event per day maximum.

CO-CHAIR ANDERSON: But, in theory, if they got transfusions on eight different days, that patient would get a numerator eight, but at this other hospital where we use revenue center codes, they could get 40 units and it would count as one, is that a correct understanding?

CO-CHAIR ANDERSON: Frank?

DR. MADDUX: So, I would say to Lorien's question, it brings up again the issue

that many, many hospitals are, in fact, do have transfusion policies and protocols that will be followed for the transfusions occurring in those hospitals.

So, the ability of the measure to be sensitive to what the expected rates are will be highly influenced by these particular areas that I think, fundamentally, puts into question some of the underlying nature of responsibility for this.

CO-CHAIR ANDERSON: All right, let's call for the vote for specifications and reliability.

MS. OGUNGBEMI: The committee is now voting on reliability for Measure 2699. The options are one-high, two-moderate, three-low and four-insufficient. Voting is open.

The results are zero votes high, seven votes moderate, 13 votes low and three votes for insufficient.

The measure does not pass on reliability.

CO-CHAIR ANDERSON: Are there any other comments for the developer on the rest of the criteria? Okay.

MS. SAMPSEL: Right, so as we have in the past, you know, the developer now has, and this will go through public comment.

But the developer could also bring additional information back to you when it comes to providing additional information on this measure for consideration.

Are there any additional insights or recommendations you would like to provide to the developers?

DR. DALRYMPLE: So, my recommendation would be to consider doing the empirical testing to see if what you observed changes if you treat procedure codes the same as revenue centers since you have the data and could just empirically compare those two.

DR. MESSANA: Okay, thank you.

DR. KRISHNAN: Maybe to follow-up on that, if there's regional variances like Frank

was saying, because regional variances may be due to hospital level or physician level, transfusion practices as well as the flavor that they're using for coding the transfusions.

DR. DALRYMPLE: So, is there anything we could recommend as a committee that would help reassure us that this isn't all being drive by hospital practices and facilities warrant accountability because we're not managing anemia correctly?

Because I think that's one of the concerns, right? We want to make sure this reflects quality of care to dialysis facility, not transfusion practices and behaviors at a hospital level.

So, is there any analysis that would help us understand whether that's what's occurring? I think that's what you're proposing, but is there a way to look at that?

CO-CHAIR CROOKS: And my comment is to go back and really examine what kind of measure is this? If it seems that it probably is not an

outcome, and I like Louis Diamond's definition that an outcome is something that happens to a patient.

I haven't really thought the subject through, but that seems to make sense and I don't know if it's a process metric or an intermediate, but I think that needs to be thought through and clearly presented so that the committee doesn't end up having to try to decide what kind of a measure it is.

DR. MESSANA: So, I'll respond to that comment, Peter.

We spent months trying to figure out whether this was cubbyholed as an outcome measure or as a process measure. Using Dr. Diamond's suggestion that something that happens to a patient, transfusions happen to patients.

So, from one perspective, some of the members of our group felt strongly that it was an outcome measure, while others thought of it as a process measure.

There's a great deal of uncertainty on

our part about what defines whether this fits a definition of an outcome or a process measure.

So, I certainly appreciate and understand the multitude of opinions around the table about that.

MS. HARTWELL: I just had a quick comment and just looking at it from the patient's perspective, is there ever a way there could be a change in policy where patients could get a blood transfusion in a clinic?

Because it's a very big inconvenience to go to the hospital to get a clinic to get a transfusion and have it be reimbursed outside the bundle. But just a thought.

DR. MESSANA: So, Lori, maybe we could -- you could ask me that -- my opinion about that offline but you've got a lot of other things to do. So, I'll defer for a time.

CO-CHAIR CROOKS: So, we're going to pause -- another pause that refreshes. You are now going to be randomly selected for a two or three year term with this committee, if you still

want to serve. No, that's not an option, sorry.

And Poonam is going to do the process.

MS. BAL: As the process, whenever you pick out a number, please into the microphone say your name and the number that you received for your term. It's going to be a two or a three.

MS. HARTWELL: Number three. Oh, Lori Hartwell, number three. So, what does this mean? Oh, okay, I guess that's -- what numbers are in there?

DR. MADDUX: This is Frank Maddux, two years.

DR. GREENSTEIN: Stu Greenstein, three years.

MS. OGUNGBEMI: I mean I'm writing it down but the transcript's also getting it.

DR. KASKEL: Rick Kaskel, three years. Does this count if you're retired?

CO-CHAIR CROOKS: As an individual, not just for a --

DR. SOMERS: Michael Somers, three.

DR. BHAN: Ishir Bhan, three.

DR. DALRYMPLE: Lorien Dalrymple, two.

MS. PAVLINAC: Jessie Pavlinac, two.

MS. LENNING: Karilynne Lenning, three.

DR. NARVA: Andy Narva, life without parole. No, three.

DR. WAGNER: John Wagner, two.

DR. STEIN: Dodie Stein, two.

MS. WAGER: Bobbi Wager, two.

DR. HAIN: Debbie Hain, two.

MS. EVANS: Beth Evans, two.

DR. FISCHER: Michael Fischer, two.

DR. KLIGER: Alan Kliger, one. No, three.

DR. KRISHNAN: Mahesh Krishnan, three.

DR. LATTS: Lisa Latts, two.

DR. KLEINPETER: Myra Kleinpeter, three.

CO-CHAIR CROOKS: Peter Crooks, three.

CO-CHAIR ANDERSON: Connie Anderson, three.

DR. ZARITSKY: Joshua, two.

CO-CHAIR CROOKS: Okay. Next up is

2701.

DR. JONES: Thank you from KCQA in particular for moving the schedule up. We appreciate that.

KCQA is pleased to review our two fluid management measures with the committee.

2701 is one of the two fluid management measures developed and tested by KCQA. 2701 measures the percentage of adult in-center hemodialysis patients in the facility whose average UFR is greater than or equal to 13 ml/kg per hour. In this case, the lower score is a better performance.

Facilities succeed on the measure and are not counted in the numerator if the dialyzed patients at an average of UFR less than 13 and/or dialyzed patients for an average of greater than 240 minutes per session.

I would like to focus on a couple of the pre-meeting committee comments.

One reviewer noted that we did not present any information on the reliability of

data source used in the testing. That is, the data warehouses of three large dialysis organizations that participated in the testing.

With the exception of patient weights, all necessary data elements for the measures are collected via what NQF refers to as an authoritative source currently with care delivery. Specifically, data are entered electronically and are not abstracted, coded or transcribed by any other person.

Because the patient weight data in some instances require transactions from chair side to the electronic system, we conducted and reported reliability testing at the level of the measure sources, which, again, is consistent with NQF.

The data are the same as those batched submitted to CMS for CROWNWeb.

There were questions about the time component in the numerator which is one of the major differences between KCQA and CMS's measure.

KCQA considers the time component a

critical element. Rather dictating the UFR remain at or below 13, the length of the session component of the measure allows judicious use of UF rates above 13 as long as the patient is dialyzed for more than 240 minutes.

With a measure to focus exclusively on the UFR without taking time into consideration, some patients may require significantly longer dialysis treatments beyond four hours, increasing the chance that they may refuse. The time component is also necessary to avoid potential adverse unintended consequences of implementing the measure on other patients who follow that patient.

In other words, if they have to go more than four hours, it might bump into the next patient schedule. Only perspective testing can address the frequency with which this scenario can occur.

Lastly, I would like to address the other major difference between KCQA and CMS measure, which is that our measure is specific to

the average rate for the session during the week that Kt/V is measured and the CMS measure relies on data for a single session.

We believe the average is more accurate and voids potential gaming. We also believe relying on a single data point is a threat to validity.

We provided the committee with additional data demonstrating that one of the LDOs analyzed and confirmed that facilities who have a Monday/Tuesday, Kt/V cycle perform more poorly than those who have Wednesday/Thursday Kt/V cycle.

Thank you for considering this.

CO-CHAIR CROOKS: Thank you.

All right, our discussants for this measure are John Wagner and Michael Somers. Who's going to kick it off? Michael?

DR. SOMERS: I can, yes.

So, in terms of evidence, we are provided with a KDOQI Guideline, which is a workgroup consensus opinion regarding the need

for the patient to be euvolemic and it does mention that some patients need to have extended HD times and slower UF. There's no real evidence in this guideline regarding outcomes and why this is ---- as a process or intermediate outcome is worthwhile.

There's also literature review provided. Three of the pieces are opinion pieces. One's data from a registry, nine are cohort or analysis of large data samples from cohort studies. That's the evidence.

CO-CHAIR CROOKS: I should mention that three committee members are precluded from discussing or voting, Constance Anderson, Lori Hartwell and Mahesh Krishnan.

Okay, John?

DR. WAGNER: Right. I just would point out that the measure requires either having 240 minutes and if one does not have 240 minutes, one needs to have an ultrafiltration rate that is less than the threshold.

And none of the articles addressed

that kind of a combination, why one 240 minute time threshold would be equivalent to an ultrafiltration rate that is whatever one wishes to define.

And although many of -- some of the articles -- at least I guess the modality of the articles mention four hours as a cut point. Some of the articles mention different timeframes as cut points and other articles mention different ultrafiltration rates as cut points for risk.

CO-CHAIR CROOKS: Okay. The floor is open for other comments on the evidence. You've got it.

MS. EVANS: I just have a question. On the UFR rate, is it greater than 13 ml/kg per hour or is it greater to or equal to -- it's conflicting in that preceding comment, the data on that?

And four hours, is it over four hours or is it over and equal to four hour dialysis time?

CO-CHAIR CROOKS: Would you like to

respond?

DR. JONES: Yes, the UFR is greater than and equal to. The time is more than 240 minutes or more.

CO-CHAIR CROOKS: Okay, Lorien?

DR. DALRYMPLE: And I just have a question for the workgroup and reviewers.

In the end, what was your overall sense of the evidence as it relates to this measure? What would you have --

DR. WAGNER: I think there are evidence that this associates to increased time with better outcomes. Again, whether it's 240 minutes or not is I think a little bit vague in the literature.

And as far as what the actual ultrafiltration rate that should be targeted as the best rate above which one does not wish to go, I think it's unclear. So I think the evidence there is a little bit weaker about what the best rate is.

CO-CHAIR CROOKS: Yes, my impression

was -- I was in a workgroup ---- that there was no systematic review presented or meta-analysis, but there was quite a few articles and 13, I think, and the preponderances and some way or another, you know, high UF bad, short dialysis bad.

And the KDOQI committee in 2006 came up with -- considered it Grade A evidence.

DR. WAGNER: Yes, I think though -- that guideline really speaks to achieving the dry weight ---- the euvolemic weight in avoiding hypertension.

So that really -- while it supports an effort to define how you get there, it really doesn't -- that guideline doesn't speak to how you get there. And of course, this guideline -- this metric does not attempt to define what the target weight is by any particular standard.

CO-CHAIR CROOKS: Yes, so the evidence isn't as strong as we'd like. It's begging out for clinical trials and I don't know if that'll ever be done or not, but evidence being what it

is, I think it's reasonably strong.

Oh, I'm sorry, Frank?

DR. MADDUX: So I'd just say that as we've moved some of the dose of dialysis measures into reserve status, this kind of measure, I think, begins to breakdown in a more granular nature some of the issues that are components of what the original Kt/V was.

And whether it's time, ultrafiltration rate or the impact of intradialytic weight gain, I think all of those become a rationale of which there's a pretty large body of literature that you could extend through the transitive property to say is related to this particular topic.

And it strikes me that there's not nearly as much known about this because we've been concentrating on the more global measure at this point and I think we have to recognize that it's not going to be quite as mature as some of the other measures.

CO-CHAIR CROOKS: Okay. Other considerations about the evidence -- body of

evidence?

Okay, we're ready to vote then.

MS. OGUNGBEMI: The committee's now voting on Measure 2701 and its evidence.

The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

CO-CHAIR CROOKS: Mahesh, I see a voting stick in your -- oh, thank you. You're having a conflict of interest.

MS. OGUNGBEMI: So, we will only have 20 votes since three committee members recused themselves.

The results are zero votes high, 17 votes moderate, two votes low and one vote insufficient.

Measure 2701 passes on evidence.

CO-CHAIR CROOKS: Okay, thank you. Gap?

DR. SOMERS: So for gap we're given data from 4,252 hemo facilities, over 412,000 patients, shows that from this group a median of 10.8 percent would not have met this measure with

a range from zero to 50 percent.

CO-CHAIR CROOKS: John, any other comments?

DR. WAGNER: No.

CO-CHAIR CROOKS: Okay. All right, well, let's vote on the gap then if nobody wants to say anything further.

MS. OGUNGBEMI: The committee is now voting on performance gap for Measure 2701. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

The results are three votes high, 16 votes moderate, one vote low and zero votes insufficient.

Measure 2701 passes on performance gap.

CO-CHAIR CROOKS: Okay, thank you. Now specifications and reliability.

DR. SOMERS: So for specifications, it's specified for CROWNWeb and all the data elements in the directions for data capture was included. It's not risk-adjusted.

In terms of reliability testing, there's data, again, given from the 4,252 dialysis facilities. They looked at the facility and treatment month as independent variables and they measured the score as variable intraclass correlation coefficient between 0.6 and 0.7 for the three LDOs which fell into a level of -- a good level of reliability.

CO-CHAIR CROOKS: John?

DR. WAGNER: Nothing to add.

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

DR. DALRYMPLE: Can I just clarify one point on this? So, was the reliability testing done with the LDOs all sending their data or was it done on CROWNWeb?

Because I thought one of the comments I saw was, does CROWNWeb have three measures of weights each month? And I don't know CROWNWeb well enough to know. So I don't know if other committee members have more insight into CROWNWeb and what data's collected monthly and then maybe

the primary reviewers just clarify on how the reliability testing was done versus the stats?

DR. WAGNER: Yes, so I didn't know if this was under feasibility, but this is obviously a data set that is generated by proprietary databases and it is assumed that CROWNWeb can be configured to provide the same data.

So, the data will be from the week that the Kt/V is calculated in that. It could be three or four treatments during that week. And so, I don't think CROWNWeb currently captures that information, so it would have to be put into the system and software developed around it.

CO-CHAIR CROOKS: Is that correct?

DR. MCGONIGAL: Yes, that is correct. We collected it through the LDOs. It's not proprietary databases, they were data warehouses, same information that's sent to CROWNWeb.

At this point, CROWNWeb is collecting one data point and so that would need to be expanded to the three.

And I think that's all you asked. Did

I miss something? Okay.

CO-CHAIR CROOKS: That's it. Okay.

DR. MCGONIGAL: Oh, yes, one other thing, we exclude four or greater treatments per month. So it would only be three maximum.

DR. DALRYMPLE: I think in the specs it says greater than four are excluded. So if it's four or more we may just need to be -- and some of the exclusions it mostly says greater than four.

DR. MCGONIGAL: Four or more is the exclusion.

CO-CHAIR CROOKS: Right. So, they're stipulating that they meant --

DR. DALRYMPLE: Details, but we're going to limit to thrice weekly? Okay.

Does anyone know if this has been proposed to CROWNWeb and would be feasible in the near future to have the three?

DR. JONES: Yes, it has been. There's some discussions about, you know, picking up on the three treatments.

CO-CHAIR CROOKS: Okay. Other comments? Okay, let's vote on reliability.

MS. OGUNGBEMI: The committee is now voting on reliability for Measure 2701. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

The results are unanimous, zero votes high, 19 votes moderate, zero votes low and zero votes insufficient. Measure 2701 passes on reliability.

CO-CHAIR CROOKS: Thank you. On to validity.

DR. SOMERS: In terms of validity, the facility-specific scores were compared to SHR and SMR. The correlations were in the preoperative direction and statistically significant.

CO-CHAIR CROOKS: John?

DR. WAGNER: I guess the only question I would have is if the SHR and SMR are going to be under review and we don't know what the outcome of those reviews are, is it fair to have validity testing then that relies on those

measures or is that a moot point?

CO-CHAIR CROOKS: Well, they're not under review yet and we still --

MS. SAMPSEL: They're technically endorsed measures.

CO-CHAIR CROOKS: Yes.

MS. SAMPSEL: And in current use.

CO-CHAIR CROOKS: So I don't think we need to worry about that. I'd just like to mention they also claim face validity with this KCQA Steering Committee and testing workgroup, for what it's worth.

Okay, any other comments on validity? Let's vote.

MS. OGUNGBEMI: The committee is now voting on validity for Measure 2701. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

The results are two votes high, 16 votes moderate, one vote low and zero votes insufficient. Measure 2701 passes on validity.

CO-CHAIR CROOKS: Okay. Thank you.

Feasibility?

DR. SOMERS: Well, in terms of both feasibility and usability, the comments that have been made already about some tweaks in CROWNWeb were pertaining.

But otherwise, the workgroup thought that the data would be generated or collected during usual provision of care through EHRs.

CO-CHAIR CROOKS: John?

DR. WAGNER: So I guess it would be perhaps important to know how difficult it will be to configure CROWNWeb to capture these data and to understand. Given the issues that CROWNWeb has with capturing all of the data, whether this would be a particular problem in this case for this metric.

CO-CHAIR CROOKS: Because as it stands, there's any patient dialysis as an independent or small chain unit is not going to be captured. Right?

DR. MCGONIGAL: I'm sorry, are you asking --

CO-CHAIR CROOKS: I'm asking you that question, yes.

DR. MCGONIGAL: Yes, so I missed part of the question but you're talking about the smaller --

CO-CHAIR CROOKS: Well, the --

DR. MCGONIGAL: -- that non-batch submitters, is that correct?

CO-CHAIR CROOKS: Right. But you're not even using CROWNWeb at all. Right?

So, at the present time, when you do this measure it excludes independent dialysis units and small chains.

MS. NISHIMI: That was for the testing. We've been in conversation with CMS about adding the two extra data points so the batch submitters would batch three and the rest of the facilities would have to manually enter the additional two, as they now manually enter one.

CO-CHAIR CROOKS: So the plan is definitely to get all units under the umbrella?

MS. NISHIMI: Oh, yes. No, there was never --

CO-CHAIR CROOKS: If we can get the -- all right, good.

Other issues with feasibility?

MS. EVANS: Home hemo will be included or not?

CO-CHAIR CROOKS: Pardon me?

MS. EVANS: Home hemo?

CO-CHAIR CROOKS: Home hemo, will they be included?

MS. EVANS: Are they included or not?

DR. MCGONIGAL: This is in center.

MS. EVANS: In center, okay.

CO-CHAIR CROOKS: It's not, yes.

Okay. Any other issues on feasibility? Let's vote.

MS. OGUNGBEMI: The committee is now voting on feasibility for Measure 2701. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

The results are two votes high, 15

votes moderate, one vote low and one vote insufficient. Measure 2701 passes on feasibility.

CO-CHAIR CROOKS: Okay. Usability and use?

DR. SOMERS: I don't think we have any new comments. We've covered that I think.

CO-CHAIR CROOKS: No comments?

DR. SOMERS: Well, nothing new than what we didn't discuss with feasibility. I mean I think that with the tweaks that needed to be made, we thought that it could be usable. It's not currently being used.

CO-CHAIR CROOKS: Okay, it's not currently in use. Pay -- state plan public reporting, could be used for plan payment, QI use.

Okay. Other concerns or comments about usability and use? I see people have -- oh, Lorien?

DR. DALRYMPLE: I'm not sure, it seems like our developers want to share something with

us. Sarah, can we --

MS. SAMPSEL: No, I'm --

DR. DALRYMPLE: -- inform our conversation about usability because perhaps we don't understand, aside from CROWNWeb, if this is usable.

MS. SAMPSEL: So I mean I guess what I'm trying to do is only if there are questions specifically to the developers that would help you in your consideration, that's when we want to direct to the developers. But we're trying to move this along.

DR. DALRYMPLE: I will just ask one question with that in mind.

Is there usability outside of CROWNWeb implementing these changes?

DR. JONES: In at least one large dialysis organization, this is reported out on a monthly basis. And for internal QI purposes, both the length of dialysis and the fluid rate and the time. So at least in that one large one it's used.

CO-CHAIR CROOKS: Any other comments? Okay, let's vote, usability and use.

MS. OGUNGBEMI: The committee is now voting on usability and use for Measure 2701. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

The results are two votes high, 15 votes moderate, one vote low and one vote insufficient. Measure 2701 passes on usability and use.

CO-CHAIR CROOKS: So, so far all the criteria have been met. Now we're going to vote on submitting for recommending endorsement. Any other comments before we vote? Okay, pick up your sticks.

MS. OGUNGBEMI: The committee is now voting on Measure 2701's overall suitability for endorsement. Options are 1 yes, 2 no. Voting is open.

Results are 19 votes yes, zero votes no. Measure 2701 passes it's meeting NQF criteria for endorsement.

CO-CHAIR CROOKS: Okay, thank you. And don't go away because we're going to go next to 2702, Post-Dialysis Weight Above and Below Target submitted by KCQA.

You have the floor.

DR. JONES: Thank you. Measure 2702 is the percentage of patients with the average post-dialysis weight greater than equal to one kilogram above or below their prescribed target weight.

First, I want to emphasize that KCQA believes that this measure complements and serves as a check and balance to the measure we just spoke about.

Implemented alone, the specifications of 2701 could be met simply by underdialyzing the patient by decreasing UFR.

Concurrent implementation of this weight measure will minimize the potential for such an unintended consequences as patients underdialyzed in such manner will unlike meet their target weight.

KCQA's testing in over 4,000 facilities over 412,000 patients found the measures were both valid and reliable.

I understand from the pre-meeting comments and the workgroup call that most of the discussion centered around the identification of the one kilogram weight as the target weight window to be achieved. So, I'd like to talk about that.

Evidence suggests that improved volume control can attenuate cardiovascular issues, making a strong case for the diligent avoidance euvolemia overload.

KCQA believes that the time for performance measurement in this area is long overdue and that patients will continue to suffer adverse consequences if we wait until an ideal measure is identified.

Research also demonstrates that overly aggressive fluid removal can be detrimental. Efforts to achieve euvolemia must, therefore, neither be too aggressive nor too weak with a

goal of consistently bringing patients to an end dialysis weight falling within a narrow margin of their defined target weight to achieve optimal state.

We know one study found absolute weight is more important than percentage weight. An additional small study of 182 patients suggested adverse consequences when the target weight was missed by more than 0.3 kilograms.

Evidence identified during our testing indicated there is significant performance gaps using the one kilogram. This value was the consensus of our steering committee, measure feasibility and testing group for the purpose of this measure.

We posit the need for performance improvement and fluid management for dialysis patients and the current absence of measures addressing this important aspect of care are best served by our consensus value of one kilogram.

We further acknowledge that the ambiguity is existing in protocols for

determining the dry weight. But, again, emphasized the need for performance measures in this area.

If the committee does not believe our reliance on our expert opinion and factors, such as lack of standardized protocols are sufficient to specify this target aim, we maintain if you utilize the NQF evidence algorithm you would still end up with insufficient but exception decisions.

Thank you.

CO-CHAIR CROOKS: And once again, the same three people are excluded from participation and our discussants are Myra and John again.

It seems like you guys get called on a lot. Did you get more than two? Who's going to go first? John?

DR. WAGNER: Go first? Okay, thank you.

So the numerator is the number of patients from the denominator with the average post-dialysis weight gain equal to or greater

than one kilogram above or below the prescribed target weight during the calculation period.

And the denominator exclusions are the following patients: patients less than 18 years of age, home dialysis patients, patients in facilities less than 30 days, patients with less than seven hemodialysis treatments in a facility during a month, patients without a 2728, kidney transplant patients with a functioning graft.

And then there's an unanswered question as to how many in-center hemodialysis patients during the reporting month need to be present in order to have the facility count within this measure.

There isn't an exclusion for greater than three treatments as we heard in the other measure. I'm not sure if that was an oversight or not.

So the evidence is, again, related to that KDOQI Guideline which argues that patients should be ultrafiltered to a target optimal dry weight. It doesn't -- and the evidence level was

A.

And in addition, the developers have an expert consensus panel in the KCQA membership and they also reviewed 14 studies that looked at issues related to technology that was used to define target weight, what the intradialytic weight gain were in various populations and what happens when one tries to achieve target weight and various adverse events.

The issue here is we're not defining what euvolemia is, we're basically allowing the clinicians to decide whatever target weight they choose based on whatever criteria they use. And then taking the one kilogram plus or minus value as being a surrogate for what presumably is going to be helpful to achieving euvolemia and control of hypertension.

So I think in terms of the evidence that is presented and the actual guideline, there is a little bit of a disconnect. This is kind of like the issue with testing for whether parachutes work or not.

I mean everyone knows that you want to prescribe a target weight. That's part and parcel of the dialysis prescription. And if one is doing that, one presumably intends that target weight to be achieved. So there should be some metric around that.

But in terms of what the evidence that that result will speak to quality, particularly if there are patient-centered factors that also impact on achievement of that target weight and there's no discussion in the evidence about what the patient-centered factors will do to one's ability to change the target weight.

And because of patient-centered factors, there might be a decision to move the target weight away from what one conceives to be the euvolemic weight. So there's no evidence presented around that subject either.

DR. KLEINPETER: So basically since there's very little evidence that's being presented because it's a new measure, there is, once again, statements regarding the importance

of establishing the target weight. But we just had just the testing data provided by the developer thus far.

CO-CHAIR CROOKS: Okay. The floor is open for discussing the evidence base linking it to the measure, linking positive health outcomes to the measure.

I might comment that I'm having trouble with this issue of knowing what the dry weight is and I know we don't have that worked out yet. But it seems to me that a patient could just be set at some weight and hitting this target all the time and be way off from where they should be.

Unfortunately, there's not another measure to link it to which is -- you know, the right target weight is, but I don't know that the evidence can really link this per se to better health outcomes.

Michael?

DR. SOMERS: I know you have to pick some target, but again, I'm not aware of any

evidence saying that one kilo is better than 1.1 kilo or 0.8 kilos.

CO-CHAIR CROOKS: Who else? Franklin?

DR. MADDUX: I just have a question for the developers.

Were there any exclusions or how did you expect to deal with patients that sign off early or actually don't have some proportion of their prescribed treatment actually delivered? Were they simply added in as failures or were they treated differently?

DR. MCGONIGAL: During our testing, they were not treated differently. If they didn't achieve the target weight, they just weren't counted in the numerator.

So it was considered as a fail, part of the rationale being that, you know, increasing patient education on the importance of staying on for their prescribed times and so underlying that.

CO-CHAIR CROOKS: I'd like to hear some more thoughts from the committee if you

think the evidence really supports that this will improve health outcomes. You do, you don't? Anybody? Frank?

DR. MADDUX: I'd just simply say fluid overload hospitalizations are an enormous component of comorbidity in this population and I think trying to shine a light on those particular issues through the achievement -- both the assessment of an appropriate target weight and the achievement of that are, I think, key issues in trying to look at how you prevent that type of morbidity.

CHO-CHAIR CROOKS: So it's a window, but not a perfect window to -- other committee members? Lorien?

DR. DALRYMPLE: I think sometimes there are practical challenges.

For example, we're often -- I think sometimes there are practical challenges that doesn't necessarily negate it but often while we're challenging dry weight, we're not re-prescribing the dry weight on a daily basis, and

it's understood that, perhaps, this would make us better and each day prescribe a new dry weight and continue to challenge from there.

And then when people return from the hospital, it's not uncommon -- at least in my experience, that they come back eight to ten kilos over their dry weight and it takes us, unfortunately, time to get there. Not that that doesn't mean this isn't important, but it can be very challenging given the frequency of hospitalization in our patients.

But I appreciate the balance of these two measures combined. But I could actually see where you would do well on one and poorly on the other. So I'm trying to reconcile then what does your quality look like?

And I'd actually be curious what other -- and I know we're short on time -- what other committee members think about if we have two quality measures on volume, one you look wonderful, one you look terrible, what the summary of that meaning is.

MS. BAL: I would just say that that's really more related in competing option. And also it wouldn't really fall into evidence. But if you want to bring it up later on, please do. But, just try to focus on evidence right now.

Thank you.

CO-CHAIR CROOKS: That's what I was going to say.

All right, so I see there's not a need to discuss this much more and -- oh, Alan, I was waiting for your card to come up. Thank you. You're on.

DR. KLIGER: So just restricting ourselves to the evidence, I think John stated it well that there really is little evidence for a compelling need to have measures for volume.

But given the arbitrary way that we clinicians set the dry weight and given lack of data on this plus or minus one kilo, the evidence, unfortunately, is not there yet.

CO-CHAIR CROOKS: Michael?

DR. FISCHER: I agree. I mean I just

was reading -- I didn't review this, I was reading what KDOQI says. I mean their focus is on dry weight, but kind of what Michael said, they don't -- there's no evidence for this parameter around how close you get to it as stipulated.

You know, and I haven't reviewed all the other 13 studies, but I don't know if the reviewer or the developer had rationale for choosing ---- putting all the side dishes around determining an optimal dry weight, around kind of this boundary around which is deemed optimal performance.

CO-CHAIR CROOKS: Thank you. Other comments? Alan, your card's still up. Okay.

All right, I think it's time to vote on the evidence.

MS. OGUNGBEMI: The committee is now voting on evidence for Measure 2702. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

The results are zero votes high, three

votes moderate, seven votes low and nine votes insufficient evidence. Measure 2702 fails on evidence.

CO-CHAIR CROOKS: So we can talk about voting on an exception -- is that why you what you were raising your hand for?

So the floor is open for proposal to consider an exception at this time.

DR. DALRYMPLE: So I propose we vote on insufficient with exception given the evidence discussion.

CO-CHAIR CROOKS: Would any others like to provide rationale for, against? Lisa?

DR. LATTS: So I am -- this seems to me to be one where if -- we let the perfect be the enemy of the good. I mean it seems to me like this is a good thing to measure and we might not have the exact right numbers, but it's probably a decent ballpark.

And this seems to me one that we should get -- and I realize it's easy for me to say it because I'm not getting reimbursed based

on this. It seems to me that this is one as a companion measure to the ultrafiltration measure and it seems to be something that is really important in the scheme of dialysis.

And that for that reason, even though we don't have the perfect data to say that these are the exact right numbers, I would be in favor of moving forward with it and tweaking as we go rather than not continuing the discussion.

CO-CHAIR CROOKS: Thank you. Would anybody else like to express a viewpoint?

Okay, well, then I think we'll just -- is this a hand vote that we do?

MS. BAL: No, this is a real vote.

CO-CHAIR CROOKS: Okay.

MS. BAL: I mean machine vote.

CO-CHAIR CROOKS: Once again, grab your sticks.

MS. OGUNGBEMI: The committee is now voting on importance to measure and report for evidence with the potential exception to empirical evidence.

The options are 1 insufficient evidence with exception, and 2 no exception.

The results are ten insufficient evidence with exception and nine for no exception and that's actually gray-zoned. So I'm actually not sure -- hold on one second.

So we're going to move forward.

CO-CHAIR CROOKS: We have a majority who would like to see this continue with an exception to the evidence. So we will go on to gap.

DR. WAGNER: Okay. So the measures -- there were data from over 400,000 hemodialysis treatments across three organizations.

And the interquartile range was 14 percent suggesting that there was a potential for some intervention that might -- if one believes this is an outcome, rather something that would improve health, that this lends itself to that.

CO-CHAIR CROOKS: Okay. Other comments on the performance gap? Okay, let's vote then.

MS. OGUNGBEMI: The committee is now voting on performance gap for Measure 2702. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

The results are three votes high, 14 votes moderate, zero votes low and two votes insufficient. Measure 2702 passes on performance gap.

CO-CHAIR CROOKS: Okay, specifications and reliability?

DR. WAGNER: So specifications, as I mentioned, we don't have exclusions for more than three treatments during the week.

And also, when there's no adjustment or exclusions for patient preference, which I think is one of the major discussions that I think patients will have regarding this.

And so -- and as far as the reliability testing goes, the ANOVA analysis indicated a moderately high interclass correlation and a high ratio of between to within facility correlations.

CO-CHAIR CROOKS: Okay. Myra?

DR. KLEINPETER: Something else, it's available through CROWNWeb so it should be relatively easy to get at this point.

CO-CHAIR CROOKS: Okay.

DR. DALRYMPLE: Is this three treatments again?

DR. KLEINPETER: Yes.

DR. DALRYMPLE: So not currently available through CROWNWeb, but we would assume it would eventually be available through CROWNWeb, correct? Okay.

CO-CHAIR CROOKS: Okay. Any other comments on reliability specifications? All right, let's vote.

MS. OGUNGBEMI: The committee is now voting for Measure 2702 on reliability. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are one vote high, 13 votes moderate, five votes low and zero votes insufficient. Measure 2702 passes on

reliability.

CO-CHAIR CROOKS: Okay. Validity testing?

DR. WAGNER: So this correlates to the standardized ratios that we've alluded to before and as well as has high face validity by the assessment of the KCQA committee tasked with developing the measure.

CO-CHAIR CROOKS: Okay. Myra, any comments?

Yes, it looked like they have strong correlation with the measure with SMR and SHR too, from my review.

Jessie?

MS. PAVLINAC: I should have asked this earlier, we've been having this discussion. Is this a clinician level or a facility level? It's not --

CO-CHAIR CROOKS: It wasn't specified.

MS. PAVLINAC: -- specified and I -- sorry, I didn't --

DR. MCGONIGAL: These are all

facility.

MS. PAVLINAC: Okay, I just wanted to know who's making the hit in the pocketbook.

DR. DALRYMPLE: And was the testing done at facility level or organization level, if that makes sense?

DR. MCGONIGAL: Facility.

DR. DALRYMPLE: Okay, and then presented by organization?

DR. MCGONIGAL: Absolutely.

DR. DALRYMPLE: Is that correct?

CO-CHAIR CROOKS: All right, are we ready to vote on validity?

MS. OGUNGBEMI: The committee is now voting on validity for Measure 2702. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are one vote high, 16 votes moderate, two votes low and zero votes insufficient. Measure 2702 passes on validity.

CO-CHAIR CROOKS: On to feasibility?

DR. WAGNER: So feasibility has the

same issues that we discussed in the prior metric. CROWNWeb would have to be modified to accept the data.

CO-CHAIR CROOKS: Okay. Any additional concerns? Okay, let's vote.

MS. OGUNGBEMI: The committee is voting on feasibility for Measure 2702. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting has opened.

Results are one vote high, 14 votes moderate, three votes low and one vote insufficient. Measure 2702 passes on feasibility.

CO-CHAIR CROOKS: Next, is it in use and is it usable?

DR. WAGNER: So yes, it's not in use. It's a new measure and so -- and presumably, it could be used in public reporting and other ways.

And I guess the unintended consequences that I would be concerned about would be impact on patients who are not adhering to their target weight prescribed by their

clinicians.

CO-CHAIR CROOKS: Developer comment briefly, please?

DR. JONES: Again, it is being used in one large dialysis organization for internal quality reporting purposes.

CO-CHAIR CROOKS: Lorien?

DR. DALRYMPLE: I was just curious if we could discuss as a committee whether others had concerns about unintended consequences, such as misprescribing the dry weight to not appear to be achieving it if this was a measure of quality and if, generally, the committee thinks the benefits outweigh any potential harms or unintended consequences?

CO-CHAIR CROOKS: John, you look like you're about to speak.

DR. WAGNER: Yes, I mean I think this is a real dilemma for clinicians when one has a patient that is clearly not -- who has defined ideas about what their target weight should be.

Whether it's a weight where they're

clearly overhydrated, but they tolerate it and they want to be at that weight and one decides that that's not the dry weight, one is left with the dilemma of do you order a dry weight or do you order the weight that the patient is only going to allow you to achieve?

And you know, I think in my own practice, after struggling with that with the patient, trying to educate the patient, change the patient's mind, I would choose the weight that the patient will allow us to achieve.

And then our facility -- our quality control was that we asked the nurses to document whenever a target weight was not achieved within 0.5 kilograms so that there be some record of why that patient did not reach the target weight and which, sometimes, is not easy to discern from the record unless there is a specific note written about that.

CO-CHAIR CROOKS: Myra?

DR. KLEINPETER: One other thing is those patients who frequently sign off early and

are nowhere near their target weight, how do you handle those in terms of will those people be adversely, I guess, dropped from those facilities for noncompliance?

Or will you just have people just adjusting things, just to try and meet with what the patient says they're going to get to?

CO-CHAIR CROOKS: Well like many measures, you're not going to hit a hundred percent, I don't think. frankly.

DR. MADDUX: So I think to John's point, we routinely use nomenclature that's probably imprecise.

So we talk about dry weights, we talk about target weights, we probably never really are prescribing a true dry weight. It's very difficult to assess. We're really looking at target weights that we're trying to achieve for an individual treatment. So I think the measure might actually give us an opportunity to really become more specific in how we use that language.

The second thing I would want to point

out is really a question on use to the group or the developers. We've never actually talked about a patient level measure, and to Myra's point, in otherwise this would be a potential measure that, in fact, a patient level measure might actually apply because so much of the early sign off or the no-show rate or the I'm unwilling to go to that weight you think I should go to discussion really is a patient decision.

And I don't think we've ever really had a patient level potential measure quite like this.

CO-CHAIR CROOKS: Did you see someone had cards up and I'm not seeing them? Oh, you're saying turn on your mic, dummy. Okay.

Okay, so ---- well, Myra, you're -- oh, you're putting your card down, okay. I almost got you.

Okay, so I think we're ready to vote on use and usability.

MS. OGUNGBEMI: The committee is now voting on usability and use for Measure 2702.

The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting has opened.

The results are one high, 12 moderate, five low and four insufficient. Measure 2702 passes on usability and use.

CO-CHAIR CROOKS: Okay, so we're going to vote now on suitability for recommendation knowing that the evidence was -- consensus was not reached on making an exception and I think the staff may have some more processing to do on what exactly -- how we proceed, but we can go ahead and vote.

Assuming that the exception is upheld, would you recommend this for endorsement to the NQF?

MS. OGUNGBEMI: The committee is now voting on overall suitability for endorsement on Measure 2702. The options are 1 yes, 2 no. Voting is open.

For Measure 2702, results are 11 votes yes, eight votes no. For Measure 2702, consensus was not reached on meeting NQF criteria for

endorsement.

CO-CHAIR CROOKS: That's not a gray zone?

MS. OGUNGBEMI: So consensus was not reached and this measure would be -- we would have to look at it again at the post-comment call.

CO-CHAIR CROOKS: Okay. Thank you very much. Okay, so do we only have two left?

You're kidding? I went back to 2700, oh my goodness, all right.

Can you go to work again, Connie?

CO-CHAIR ANDERSON: No.

CO-CHAIR CROOKS: No?

CO-CHAIR ANDERSON: None of us can since it's a competing measure with the KCQA measure.

CO-CHAIR CROOKS: I'm not even finding it here.

CO-CHAIR ANDERSON: On the last page.

CO-CHAIR CROOKS: Oh, I went the wrong way. Okay, we'll consider Measure 2700. This is

ultrafiltration rate greater than 13 ml/kg per hour from U of Michigan and CMS.

Are the developers here? Yes, okay. Take it away.

DR. MESSANA: Such as we are.

So I would like to say that I love the KCQA measure about ultrafiltration and I know some in my group have commitment disorder kind of issues, and so they just are very fond of it, but I personally love it. Okay?

And the reason I love it is that our measure is very similar fundamentally to the KCQA measure that you discussed earlier. And the similarities I'd like to highlight as you debate the two measures -- debate our measure.

First, there is an approximately 13 ml/kg per hour threshold for the two measures. Second, both measures attempt to exclude transient and new patients. Third, the measures share a common evidence basis. And fourth, the two measures, even though they are somewhat different, as is pointed by KCQA, have almost

identical validity and reliability outcome scores when you look at them. So they are fundamentally very similar.

There are a couple of differences that are noteworthy. KCQA has pointed out that an additional data requirement will be needed in CROWNWeb before their measure can be operationalized, okay? At least using CROWNWeb data.

Ours does not require that. So the potential threat to validity that was raised about our measure when the comments were made needs to be weighed against the feasibility issue. Our measure is calculable right now. Okay?

However, it's not to say that we disagree with potentially using thrice weekly ultrafiltration data when it's available, if it's available and if the data collection burden for single user interface facilities is addressed and we'll see what that timeline is.

The one difference that I think is

important is the 240 minute statement in the numerator that was raised by some. Ours does not include that and we have had some difficulty trying to harmonize over that particular aspect of it.

As has been pointed out, the literature doesn't address that particular numerator statement or practical considerations that KCQA used, but we have been working on analyses looking at the 240 minute numerator inclusion or numerator statement, and find that about 20 percent of the patients ---- based on the current CROWNWeb data or recent CROWNWeb data, about 20 percent of the patients who have high UFR rate have high UFR rate with a 240 minute or longer treatment time.

And if you look at associations with mortality for those patients stratified on treatment time cut at 240 minutes, the increased rate -- the increased mortality associated with high UF rate, the association is at least as strong for patients above 240 minutes as it is

for patients below 240 minutes.

So if we get around, if we have the opportunity to harmonize measures down the road, that's going to be an important issue to talk about. But other than that, those are relatively small issues. Fundamentally, the two measures are relatively similar.

Thank you.

CO-CHAIR CROOKS: Okay, thank you.

The same three people are, I believe -- yes, excluded from consideration -- considering this measure or voting.

And our discussants are Rick Kaskel and Dodie Stein. Take it away Dodie.

DR. STEIN: According to this, it's an intermediate clinical outcome intended to guard against risk with rapid fluid removal, high UFR, faster UFR.

From my own perspective, I think the goal is a good one and it sounds like the evidence is compelling. I have some other concerns about how this thing happens at the

clinical level.

Rick?

DR. KASKEL: Just to add -- and again, basically, we're with this challenge that early data demonstrated that a high ultrafiltration rate was linked to increased mortality, increased intradialytic hypotension as well as myocardial stunting.

So it is very important data indeed and there is some data that suggests that this is mechanistic and should be evaluated.

CO-CHAIR CROOKS: Okay, Alan?

DR. KLIGER: Can I point out that all of those data which are interesting are associative and not causative data.

And patients who get those high ultrafiltration rates, particularly those who get long treatments and high ultrafiltration rates may be the very ones at highest risk of death, not because we're dialyzing them or ultrafiltering them but because they're the big weight gainers and have cardiovascular ----

underlying substantial cardiovascular disease.

So I just think we have to be very careful interpreting those data.

CO-CHAIR CROOKS: Other discussion of the evidence? Franklin?

DR. MADDUX: I'd just say that as I've looked at the evidence on ultrafiltration rates on a per kilogram per milliliter per hour, it's the -- it seems fairly obvious when you're below ten or you're above 15.

Between 13 and 15, there's still some controversy, I think, in some of the articles that I've read as to exactly where that cut point is. So it's a bit of a less clear zone and that was one of the reasons why I thought the additional time-based element to the measure had some value because it mitigated that to some degree for these patients that you mentioned, Alan, that may represent as many as 20 percent of the over 13 people.

There is a lot of study still to be done in this area, it seems to me. And so we're

not at a high level of maturity, I think, on this particular topic. So it strikes me.

And then the other is simply a question that I have probably for the group and then for the developers would be my big concern about the single measurement in this is the long intradialytic interval and the impact that that would have on the results by using that single measure if it happened to be on the Monday or Tuesday -- typical Monday or Tuesday.

And I don't know whether that bothers anybody else or not.

CO-CHAIR CROOKS: That might be reserved for the specifications section then.

So the evidence is not what we'd like, but it's the evidence that we have.

Other -- Lorien? I'm sorry, I didn't see your card.

DR. DALRYMPLE: So I just have kind of one general question for the group since we did just recently review a measure for KCQA that's very similar except for in the assessment of

time.

Does anyone have a different assessment of evidence for this measure as compared with that one? Or a different interpretation of the evidence than we did for the prior measure that had a similar target?

CO-CHAIR CROOKS: No? Okay.

Yes, Michael?

DR. SOMERS: Except that some of the evidence in the other measure was the time-based evidence.

So I think that would have been considered along with the evidence for the ultrafiltration goal. So I think that's kind of the difference between the two in my mind.

DR. DALRYMPLE: Did your impression change between these two measures with that --

CO-CHAIR CROOKS: Mic?

DR. DALRYMPLE: I'm sorry. And my question just was, you know, to the group.

I'm just curious if you would rate the evidence differently based on those differences?

And I'm just curious about the group's thoughts on that since we just recently --

DR. SOMERS: Well in my mind, I thought the time-based evidence was stronger than the UF rate evidence. And other people on my workgroup said that as well. We didn't ask everyone individually, so I don't know if there was consensus in the workgroup.

CO-CHAIR CROOKS: John?

DR. WAGNER: I guess I would add that with the double measure, there are two patient-related factors in the metric.

In here we have one, namely the patient's intradialytic weight gain can only be handled in a certain way in terms of UFR. And if a patient is willing to extend time to at least the four hour mark in the other metric, then whatever intradialytic weight gain the patient has falls away as an issue.

So in this case, we don't have that as a safety valve for the patient in any case.

CO-CHAIR CROOKS: Okay. Shall we vote

on the evidence?

MS. OGUNGBEMI: The committee is now voting on evidence for Measure 2700. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

MS. BAL: Could everyone try one more time? We need one more vote. There should be 19.

Unless someone -- unless you see your friend missing other than -- oh okay, then we're good. Never mind. It's hard to see on this side.

MS. OGUNGBEMI: Well, it's a gray zone measure because the results are zero votes high, nine votes moderate, six votes low and three votes insufficient. So Measure 2700 is in the gray zone for evidence.

CO-CHAIR CROOKS: Okay. Familiar territory for this group.

Okay, so we're going to go on to discuss the performance gap.

DR. KASKEL: I'll do it.

So this is CROWNWeb analysis of over 400,000 hemodialysis entries and it shows a ten percent splay in performance between the 25th percentile and the 75th percentile facilities.

Given the number of HG patients nationally, it translates to a large patient population with a potential benefit if the rationale behind the measure holds.

There's disparity data presented that looks at key facility demographics separated into quintiles, all of which showed differences in performance across quintiles. However, it's not clear if this is biologically a basis for these differences or changes that might result in care.

So the gap information would benefit from the national measure and their suggestion that there is disparity data differences.

CO-CHAIR CROOKS: In my interpretation of the disparity data was that there was statistically significant data because they have so many numbers to crunch, but they may not be clinically meaningful, you know, the differences

between the quintiles.

Okay, let's vote on performance gap then.

MS. OGUNGBEMI: The committee is now voting on performance gap for Measure 2700. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting has opened.

Results are zero votes high, 17 votes moderate, one vote low and one zero votes insufficient. Measure 2700 passes on performance gap.

CO-CHAIR CROOKS: Okay, so we can continue on our discussion about specifications and reliability.

DR. STEIN: Yes, on the reliability, 84 percent variation was attributed between facility differences and 16 percent attributed with in-facility variation.

Testing was performed on adults hemodialysis patients at dialysis facilities over 11 patients. CROWNWeb data overall IUR was 0.84, as I said.

CO-CHAIR CROOKS: Okay. So the IUR was good and we were discussing specifications a while back.

Franklin, anything else you wanted to underline or make clear?

DR. MADDUX: No, I think what I said before is where I am. I'm still concerned about the long intradialytic interval from the one measurement versus several.

CO-CHAIR CROOKS: Okay. Other concerns about the specifications or reliability that you'd like to mention?

Okay, let's vote.

MS. OGUNGBEMI: The committee is now voting on reliability for Measure 2700. Options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting has opened.

Measure 2700 for reliability has passed. The votes are zero high, 12 moderate, five low and one insufficient.

CO-CHAIR CROOKS: Yes, 75 percent.

Okay, what's next? I've lost my

place. Validity.

DR. KASKEL: So validity. The analysis provided showing that there can be a clear separation in measurement performance to group facilities.

Given that there was less compelling data to show interquintile validity differences, this would need to be -- one would need to be careful how such a construct was designed to stratify centers to make sure that potentially non-significant differences in quintile distribution were not part of the design.

Some units determine kilograms to be removed first then time-allowed resulting in ml/kg per hour measure. What needs to be established and controlled first is the question. What effect would standard would have on patients' comfort -- the standard have on the comfort, adherence to treatment, feasibility of dialyzing patients in a typical four hour window, some units use three to three point five hours, and the impact of staff effort with the patient

needing attention for symptoms such as cramping, blood pressure and the costs ---- the unit costs.

These are a lot of significant areas here for testing. Should standard specify limits on fluid removal per hour as well? And should it consider mortality rate of the unit relative to ultrafiltration?

Some units may have more intradialytic weight issues and treatment non-adherence as a result of the patient mix and during the dialysis treatments would be different with the problems. And what about the socioeconomic or psychosocial issues? All of these things have to be taken into account.

You had some other discussion about patient involvement or decision?

DR. STEIN: Yes, I'm not sure. This may be feasibility, but just it takes the choice out of the patients. It's not patient-centered and there are a number of other issues that I'll mention under feasibility that go along with this.

CO-CHAIR CROOKS: Okay. Let's just stay with the validity question for the moment and then we'll -- can open up these other issues.

So any other questions on validity? Lorien?

DR. DALRYMPLE: So I just wasn't sure if we wanted to discuss as a committee the findings from the Poisson regression on the SHR and SMR with respect to the highest quintile.

There is some explanation by the developer, but the lack of graded association and what other people make of that.

CO-CHAIR CROOKS: She's saying the validity testing gave negative results or --

DR. DALRYMPLE: So it didn't give graded results and there is some explanation of that.

So for example, quintile five, the relative risk for the ---- let me make sure I'm doing the hospitalizations is for quintile five, 1.03, but for quintile two, 1.05, for three 1.08, for quintile four 1.06.

So in other words, the highest rates of UF have relatively lower hospitalization or mortality. And I think the developer's fairly proposed mechanisms why that may be, you know, are patients that tolerate very high UF rates tend to be younger, perhaps healthier at baseline.

But I just didn't know if the committee wanted to discuss, since this was the validity testing put before us, if it influenced out thoughts on UF rate and outcomes.

So there's also mortality on my document, this is page 25. I don't know if Poonam or someone could just pull up page 25 and it's a little bit easier if you see it than speak it.

CO-CHAIR CROOKS: I'm one measure behind.

DR. DALRYMPLE: Page 25 on my worksheet, which I think includes comments.

And the developers will correct me if I'm wrong, but they say this is likely a

manifestation of selection bias relating to healthier patients being more tolerant of higher UF rates. Facilities in the highest quintile may, therefore, have greater proportions of healthier patients in their panel, blunting the current risk of higher UF rates.

Adjustment for this effect would require a complex analyses beyond the scope of what is possible here, which I understand those would be fairly complex. I just didn't know if this affected interpretation of --

CO-CHAIR CROOKS: Before we go to the developers, are there other responses?

I have to say, it doesn't reassure me that's a valid measure from this analysis.

Alan?

DR. KLIGER: Yes, great question. So the alternate explanation other than what the developer has is that the underlying hypothesis is incorrect.

The other possibility is that these higher ultrafiltration rates don't result in

worse outcomes, in fact. Now we don't know because the data all are associative data and not causative data. And so, I think you're observation is a very important one.

DR. FISCHER: Did they provide anything else for validity testing other than this?

Because if this is it, then if you once again go back to the NQF algorithms, this falls into kind of a low category in terms of validity.

But I didn't review this. I don't know if there's anything else in the application beyond this commenting on face validity, et cetera.

CO-CHAIR CROOKS: Yes, I was in this workgroup and I share that concern that the validity hasn't been established.

DR. DALRYMPLE: Can I ask just a procedural question? Because our committee just, I think two measures ago, voted to pass a very similar measure although the specifications are

different.

This is one value versus three values, which may be part of why there's a disconnect here. We could argue there's a specification difference.

But, Sarah, do you have insight into when committees have this inconsistency between similar measures and how we assess validity?

MS. SAMPSEL: Well, I mean in this case, you also -- I mean I assume the testing was different on the other measure. So that's the other difference here.

You know, we'll still bring you back to the point, you know, now you need to vote on what was provided here and then during public comment, et cetera, you know, if the developers have additional information for your consideration as well as through public comment have additional information for consideration, that would be brought back to you.

But I mean you are correct. That's a really good question because there should be

consistency in the voting.

I feel that you all did talk about on the evidence side, you know, why that you voted differently on the evidence. And so, this brings out why you may be voting differently on validity.

CO-CHAIR CROOKS: Any other comments? Questions? John?

DR. WAGNER: So I'm just curious, is this really an insufficient or low validity outcome or is this a no?

CO-CHAIR CROOKS: A no validity outcome?

DR. WAGNER: Yes.

CO-CHAIR CROOKS: I guess that's your choice exactly how to categorize it, but we need to do it now, so let's raise our sticks and do our job.

MS. OGUNGBEMI: The committee is now voting on validity for Measure 2700. The options are 1 high, 2 moderate, 3 low, 4 insufficient. Voting is open.

The results are zero votes high, four votes moderate, eight votes low and seven votes insufficient. Measure 2700 fails on validity.

CO-CHAIR CROOKS: Now that's not the end of consideration for validity -- oh, wait, this is validity, it is, yes.

Okay, so we can stop considering this measure. We'd like to offer a chance to give feedback to the developers before we move on.

DR. DALRYMPLE: My only question would be, I know in some of the prior measures where there's been inconsistent findings, measure stewards have provided us with face validity which we have credence to.

So if there is a TEP or someone else who's weighed in that could grant face validity?

DR. FISCHER: My comment would be to look at your own data and maybe rethink alternative hypotheses and explanations.

And maybe one of them may be the specification issue that Franklin and other people raised, having a one-time value and maybe

that explains why in the validity testing you have an unexpected finding that may not otherwise appear if it's specified differently.

CO-CHAIR CROOKS: Okay.

MS. WAGER: I'd like to make a comment. I think with this measure, the focus is not, as Dodie mentioned, the patient-centered care.

When I worked on the dialysis floor, we always assessed the patient individually. Their goal was individual. I think this -- you're taking the patient out of the equation when they are the most, I think, important person of that team at that time when they're dialyzing.

If we don't consider them or take what their ideas like Dr. Wagner said, we could have adverse events. I'm sorry, I just think the patient should be involved in it.

CO-CHAIR CROOKS: Very good, thank you. Any other comments for the developers? Advice? Dodie?

DR. STEIN: I continue to be concerned

about the patient and how you juggle all these measures without the patient having any choice in the matter and how one juggles a limited time, a specified time, plus the dry weight or whatever the goal is with that, plus an ultrafiltration rate.

And it just seems very complex and burdensome on staff as well.

CO-CHAIR CROOKS: Okay. Thank you. Any other comments?

Okay, thank you very much.

DR. MESSANA: Can I ask for clarification?

CO-CHAIR CROOKS: Yes, you may.

DR. MESSANA: So regarding the patient participation comments, I'm having a little bit of trouble understanding how this metric is different in that regard and that concern than the metric you passed two measures ago. Thank you, thanks -- that's the M word.

So I'm just trying to better understand that because I really -- you know, I

value the perspective, but I want to understand it better.

MS. WAGER: Again, I'm just one of 23.

CO-CHAIR CROOKS: Okay, so moving forward, we are going to -- a couple of points.

In 20 minutes, we're going to stop for public comments. The staff has advised me that they want to continue to work as long as we have a quorum, which is 15.

So after 3:00 if you need to go, you need to go but we'll try to get a little -- squeeze a little more work out of today.

We have a two hour call scheduled next week.

MS. BAL: Next Tuesday.

CO-CHAIR CROOKS: Next Tuesday. Okay. Did I cover it? Okay.

I'm turning it over to Connie for the next measure.

CO-CHAIR ANDERSON: All right, we're going back to Measure 0225 which is the measure of serum phosphorus concentration. And Claudia

and Joel? 0255, what did I say?

MS. BAL: 0225.

CO-CHAIR ANDERSON: Oh, I'm sorry. It's really hard to sit for an hour and a half and keep your mouth shut. This is really unfair.

All right.

DR. DAHLERUS: Okay, so yes, this is Measure 0255 and this is a process measure of monthly phosphorus measurement. It was originally developed in 2006 and endorsed by NQF in 2007 and again in 2011.

A mineral bone disorder TEP was convened in 2013 and among their activities was reviewing the current process measure to also consider expansion of the measure to include serum or plasma, as well as another review of any additional evidence.

The TEP affirmed the measure. They did propose a revision to the denominator to include children in the denominator feeling it was important that children are assessed for a monthly measurement.

And then they also wanted to ensure that transplant patients with a delayed graft use ---- a functioning graft, but delayed graft use, would still be included in the monthly assessment of phosphorus feeling that it was important if they were on chronic dialysis they would be included.

CO-CHAIR ANDERSON: All right, and the discussants are Ishir and Michael.

DR. BHAN: So I'll start this off. So just going quickly.

So the evidence here, there is a couple of sources provided, but one is the KDOQI Guidelines -- sorry, these are the KDIGO Guidelines, that suggest for CKD stage 5D or stage 5 including 5D that phosphorus be measured every one to three months.

Then there is a whole bunch of other references provided which are not summarized per se, but I know from my own experience, there is a reasonable associative retrospective data on high phosphorus levels and increased mortality, though

that's associative data so there is always limits to that.

Main comments, this is, of course, focused on the -- it's a process measure, so it's focused on the measurement of phosphate. This data was not directly about the measurement of phosphate, it was largely on the associations of phosphate with outcomes. Of course, if you don't know what the phosphate is, then you wouldn't be able to act on it.

That said, there is a little bit of discrepancy here. The KDIGO Guidelines say one to three months and this measure does state a monthly phosphorus, so there is a little discrepancy of note there.

Michael, did you have anything to add?

DR. FISCHER: I agree with all that. I think we struggled -- and Ishir, please chime in, and our other workgroup members, about evidence. And this was a measure that was endorsed, I think, before and is up for reendorsement.

I think at that point, we also struggled a bit ---- I mean I'm almost of two minds about process measures. On one hand, they're tightly linked, obviously, to provider behavior and they're infinitely measurable. On the other hand, there are so many steps between them and what we really care about in outcome.

And I think that, to me, always is a struggle in terms of evidence. And I think Ishir nicely pointed out, a lot of the evidence is about actually phosphorus levels, not necessarily the act of measuring phosphorus, and that's what this is about.

I think the only thing I would add is, this was reviewed by KDOQI, but my understanding -- and you can correct me, is that it was without grading. And I don't know what without -- and maybe others on the committee know better than me, I don't know what without grading is. Is that D plus?

So you know, I'm not really sure how to interpret that, but it certainly is not a high

level of evidence. I think otherwise I agree with all the comments and if other people from the workgroup have additional recollections, it'd be good to hear from them.

CO-CHAIR ANDERSON: Open for discussion by the committee.

No discussion, then let's call for the vote on evidence.

MS. OGUNGBEMI: The committee is now voting on evidence for Measure 0255. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are one vote for high, 14 votes moderate, two votes low and four votes insufficient. Measure 0255 passes on evidence.

CO-CHAIR ANDERSON: Moving to performance gap?

DR. BHAN: So for performance gap here, the data was generated from the 2013 CROWNWeb over 6,000 facilities.

And the key numbers here are the quartiles, the 25th percentile was 86 percent,

the 50th percentile, 92 percent and 75th, 96 percent. So if you look at the median 92 percent, that doesn't leave a whole lot of room for improvement.

I also wondered out loud without any real answers as to why people aren't having phosphorus measured, but I supposed that's, you know, the same issue came up with the hemoglobins in the pediatric population. But we don't have answers here about that.

So I think that was the real concern that came up on our call was how much room for improvement is there? Are there factors that can actually be intervened upon? Is this hospitalization-related issues, et cetera?

CO-CHAIR ANDERSON: Alan?

DR. KLIGER: Just to point out that a patient would fail this if they missed one month. Right? So if they were in the hospital or somewhere else or -- so, it's not a surprise.

DR. BHAN: Although they're an exclusion here. The denominator excludes

patients who have not been in the facility the entire reporting month.

Although that may be more transfer-related than hospitalization. So you might have a valid point. It also speaks to the discrepancy of the one monthly versus between one to three months.

DR. FISCHER: I mean I think over all, is it fair to say I think we felt the performance gap was rather small, is that fair? I think that was kind of the consensus of our working group. Just so everyone has an idea what the -- I think the working group sentiment was at the end of our conversation.

CO-CHAIR ANDERSON: Frank?

DR. KRISHNAN: I think it's currently measured in the QIP, right? So it's small in the QIP as well, the calcium phosphorus measurement is a QIP measure today is small, the gap.

CO-CHAIR ANDERSON: Frank?

DR. MADDUX: Could I just -- can I get a clarification?

So the denominator exclusion is patients who haven't been in the facility for the entire month, so it would exclude anyone who is hospitalized. Is that correct? Yes, that's the way I interpreted it.

DR. BHAN: Can we confirm that with the developer?

DR. DAHLERUS: So if the patient was hospitalized for the entire month, they would be excluded. If they were hospitalized for a part of the month, they would not be excluded.

DR. MADDUX: So that isn't quite how I read what's on here. It says exclusions are implicit in the denominator definition and include all patients who have not been in the facility the entire reporting month.

DR. DAHLERUS: So the entire reporting month is meant to be 30 calendar days. So they have to --

DR. MADDUX: So if I was out for five calendar days, I wouldn't have been there the entire reporting month?

DR. DAHLERUS: Right, but you have to be out of the facility, discharged from the facility for the entire 30 days. If not, then you would still be included.

DR. MADDUX: Okay.

DR. DAHLERUS: And we'll confirm with --

DR. MADDUX: It's a little -- it's not quite clear --

DR. DAHLERUS: Right, we'll confirm.

DR. MADDUX: -- the way it's written.

DR. DAHLERUS: Agreed.

CO-CHAIR ANDERSON: Any further discussion? All right, let's vote on performance gap.

MS. OGUNGBEMI: The committee is now voting on performance gap for Measure 0255. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are zero votes high, six votes moderate, 14 votes low and one vote insufficient. Measure 0255 fails on performance gap.

MS. SAMPSEL: So, I muted myself, but this is ---- for consistency sake, this would be very similar to other ones where ---- that you considered for reserve status.

CO-CHAIR CROOKS: Right, right, right. Thank you.

CO-CHAIR ANDERSON: I guess we're going to do a hand vote. Would the committee like to consider making this measure ---- bringing to reserve status?

DR. KRISHNAN: Just a question for Sarah, though. This is not currently a fielded metric, right? It's a component of -- calcium phosphorus is a fielded metric, but phosphorus itself is not.

DR. DAHLERUS: It is a current measure, yes.

DR. KRISHNAN: Calcium and phosphorus is a current measure. Phosphorus itself?

DR. DAHLERUS: No. This is the serum phosphorus is a separate current NQF endorsed measure.

DR. FISCHER: Right, and again, this was endorsed originally in 2007, re-endorsed in 2012.

CO-CHAIR ANDERSON: Would the committee like to vote this for reserve status? All right, hands if you would like this to move to reserve status?

All right, we'll continue with reliability.

DR. BHAN: So for reliability, I'm just going to jump to our comments.

The specifications, there was some confusion about inclusion of transplant patients with an active allograft. There was also concern about inclusion of pediatric patients since the evidence cited was from adults. I want to hear from our pediatric colleagues about that, but there was some concern that exclusion criteria would thus need to be modified.

And the same discussion we already had about the effects of hospitalization specifically being an issue that people are hospitalized for

part of the month would still be included.

The reliability testing -- let me just pull that up, looked at IURs that were 0.95 to 0.97. Most of the variation was attributed to between facility variation and it looked overall pretty reliable.

So to summarize, some concerns about the specifications that we discussed.

CO-CHAIR ANDERSON: Michael, any addition?

DR. FISCHER: Yes, I agree with Ishir. I think we had questions for the developer about the items he mentioned.

I think we already discussed the hospitalization thing, which I found confusing. I interpreted it the way Franklin did when I read this. And then the other two issues were pediatric patients and transplant patients with functioning allografts.

CO-CHAIR ANDERSON: I think one of the other questions at least under the reliability in the specifications was about the home dialysis

patient population and I'd like to ask the developers if they've come up with what are you going to do -- are you excluding the home patients? Are they included?

DR. DAHLERUS: I believe the home patients are included.

CO-CHAIR ANDERSON: They are included?

DR. DAHLERUS: I believe -- this is something that I'll have to check with our analyst to make sure that they were included in the analysis, in the calculation.

CO-CHAIR ANDERSON: Okay. For both hemo and PD?

DR. DAHLERUS: Yes.

CO-CHAIR ANDERSON: Okay.

DR. DAHLERUS: So we do indicate that home dialysis patients are included as I'm looking here at my notes. So I'm not sure why that was a question to begin with.

DR. FISCHER: And this is including adult and pediatric patients? Just to confirm.

DR. DAHLERUS: Yes, it is. And that's

a change from the measure that was originally endorsed in 2007 and again in 2011.

The 2013 MBD TEP wanted to add children to the denominator in order -- that they would be assessed on a monthly -- for the monthly measurement as well.

DR. FISCHER: Okay, thanks.

CO-CHAIR ANDERSON: Any other comments on reliability? All right, let's vote.

MS. OGUNGBEMI: The committee is now voting on reliability for Measure 0255. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are one vote high, 14 votes moderate, three votes low and zero votes insufficient. Measure 0255 passes on reliability.

CO-CHAIR ANDERSON: All right, moving on to validity?

DR. BHAN: So validity here, we talked about the issue of the where this is coming from, primarily data around phosphorus and outcomes

rather than testing per se.

The validity testing here looked at Poisson regression models to determine the association with mortality and came up with a ten percent -- a relative risk of mortality of 0.98 for a ten percent increase in the performance score.

I guess the issue there is how many facilities can have a ten percent increase in the performance score? And even if they achieve that, there's relatively small association with mortality here.

And I think there was this -- also a question about the validity given the issue of patient absences.

CO-CHAIR ANDERSON: Michael? Any further discussion on the part of the committee?

All right, we will be voting on validity.

MS. OGUNGBEMI: The committee is now voting on validity for Measure 0255. Options are 1 high, 2 moderate, 3 low and 4 insufficient.

voting is open.

Results are one vote high, 16 votes moderate, two votes low and zero votes insufficient. Measure 0255 passes on validity.

CO-CHAIR ANDERSON: We're going to pause here for a minute. It's time for public comment. So we'll take the measure up after the public comment time.

OPERATOR: Okay. At this time, if you would like to make a comment, please press star then the number one.

There are no public comments at this time.

CO-CHAIR ANDERSON: All right, so let's move on. Feasibility?

DR. BHAN: So in general, our group felt that this is feasible.

It's already part of the QIP and data is routinely generated during the process of care. So it looked pretty feasible.

CO-CHAIR ANDERSON: Michael?

DR. FISCHER: Let's vote.

CO-CHAIR ANDERSON: All right, let's vote.

MS. OGUNGBEMI: Committee is now voting on feasibility for Measure 0255. Options are 1 high, 2 moderate, 3 low, 4 insufficient. Voting is open.

Results are 12 votes high, seven votes moderate, zero low and zero insufficient. Measure 0255 passes on feasibility.

CO-CHAIR ANDERSON: All right, moving on to usability and use?

DR. BHAN: It's used in the QIP. There's not any obvious unintended consequences publically reported. It seems useable.

CO-CHAIR ANDERSON: Alan?

DR. KLIGER: So in terms of usability, I want to ask the pediatricians their opinion. Monthly blood draws from your home dialysis patients?

DR. SOMERS: Since almost all of our home dialysis patients are PD patients, they're seen by us on a monthly basis anyway. I don't

think that's a big issue in the pediatric world.

DR. KASKEL: I concur. They come in and even if they're a small infant, we need to get -- we get the labs.

CO-CHAIR ANDERSON: All right, ready to vote on usability and use.

MS. OGUNGBEMI: The committee is now voting on usability and use for Measure 0255. Options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are 11 votes high, eight votes moderate, zero votes low and zero votes insufficient. Measure 0255 passes on usability and use.

CO-CHAIR ANDERSON: All right. We are -- any other comments or issues to discuss before we vote on usability for endorsement and for reserve status? So let's vote.

MS. OGUNGBEMI: Committee is now voting on potential for reserve status for Measure 0255. Options are 1 yes, 2 no. Voting is open.

Results are unanimous, 19 votes yes, zero no for potential for reserve status for Measure 0255.

CO-CHAIR ANDERSON: We still have a quorum and so we would like to continue if the committee is willing to continue on until we lose quorum until at least 3:00.

Okay, thanks. Next measure is 1425, Joe and Joel?

DR. MESSANA: For the sake of time, I will limit my comments on this measure to clarification of a typographical error that was --- or a question that was raised about a potential typographical error in 1(b)(2) and 1(b)(4).

In the workgroup, a question came out was it 30 facilities or 13 facilities. The answer is 30 facilities. I went back and looked at all the analyses supporting the submission and the 13 is a typographical error. It's supposed to be 30 facilities.

I will leave it at that since you're

pressed for time.

CO-CHAIR ANDERSON: All right. Andy and Debra?

DR. NARVA: Yes, this is measurement of nPCR for pediatric hemodialysis patients and it's the percentage of patient months of pediatric in-center hemodialysis patients with documented monthly PCR measurements.

It's a level analysis is the facility, the evidence is based on KDOQI Guidelines. A general nutrition guideline from 2006 and the 2008 KDOQI Guideline on Nutrition in Children with CKD, which recommended nutritional status and growth of all children with CKDs two to five including 5B should be evaluated on a periodic basis. That was given a Grade A strong evidence.

They provided data from 2014 literature view which was supportive, and in addition to that, there's some evidence that in adolescence, PRC levels were an earlier and more sensitive marker of -- than serum albumin of malnutrition.

Despite this, there's not a vast amount of data to support this measure and a lot of this is inferred from benefit in adults.

CO-CHAIR ANDERSON: Debra, any additional? Open for discussion to the committee.

All right, let's call for the vote on evidence.

MS. OGUNGBEMI: The committee is now voting on evidence for Measure 1425. Options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are two votes high, 16 votes moderate, zero votes low and one vote insufficient. Measure 1425 passes on evidence.

CO-CHAIR ANDERSON: Performance gap?

DR. NARVA: Okay. Nutrition is obviously an incredibly important indicator in all dialysis patients, particularly those that are growing or trying to grow.

2013 data from CROWNWeb showed a mean score of 80.4 percent and that included 455 in-

center hemodialysis pediatric patients from 30 dialysis facilities with at least 11 eligible pediatric patients. So there appears to be a gap.

CO-CHAIR ANDERSON: Debra? Any further discussion on the part of the committee?

All right, we are ready to vote for performance gap.

MS. OGUNGBEMI: The committee is now voting on performance gap for Measure 1425. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are two votes high, 17 votes moderate, zero votes low and zero votes insufficient. Measure 1425 passes on performance gap.

CO-CHAIR ANDERSON: Moving on to --

DR. NARVA: Reliability?

CO-CHAIR ANDERSON: Yes.

DR. NARVA: Okay. This measure is meant to be collected and calculated through CROWNWeb. It seems to be fairly clearly

delineated.

There was a comment that the exceptions to the denominator were confusing, but that wasn't from me and I'm not sure what the commenter meant exactly.

It's a process measure and it's not stratified or risk-adjusted.

There is in reliability testing, is that a separate -- I can't remember. Did we vote on that separate? Okay. Together, yes.

Well, from the 2013 CROWNWeb data, the overall IUR was 0.985, which indicates about 98.5 percent of the variation is between facility differences.

CO-CHAIR ANDERSON: Debra? Any discussion on the part of the committee?

DR. NARVA: Oh, there was one issue with validity, which is that --

CO-CHAIR CROOKS: We're not doing validity yet.

CO-CHAIR ANDERSON: We're on reliability.

DR. NARVA: Oh, sorry. Yes, okay.

CO-CHAIR ANDERSON: You're okay, yes.

DR. NARVA: I'm rushing it.

CO-CHAIR CROOKS: I just wanted to ask the pediatric nephrologists, this calls for the measurement to be done monthly. Is that really what you recommend? Things can change that fast?

DR. ZARITSKY: That's what we do, yes.

CO-CHAIR CROOKS: Okay, so you agree with that? Okay.

DR. ZARITSKY: That's the same data that's a degenerate of Kt/V so it's just added information.

CO-CHAIR CROOKS: It doesn't add a further burden to the patient really?

DR. ZARITSKY: No, I mean utility is another story, but measuring that once a month's not a problem.

CO-CHAIR ANDERSON: All right, ready to vote on reliability.

MS. OGUNGBEMI: The committee is voting on reliability for Measure 1425. The

options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are two votes high, 17 votes moderate, zero low and zero insufficient. Measure 1425 passes on reliability.

CO-CHAIR ANDERSON: Moving on to validity?

DR. NARVA: The evidence seems to support the measure is consistent except among the facilities with 11 eligible pediatric patients with recorded PCR values.

Facilities with a hundred percent reporting had a mean serum albumin of 377 while facilities with less than a hundred percent had higher albumins, which, you know, on the face of it, appears to suggest that the evidence is not consistent with -- the measure's not consistent with the evidence, but it appears likely or leads -- speculation is that the more frequent PCR determinations were done in the patients who had poorer nutritional status, hence the bias.

CO-CHAIR ANDERSON: Debra? Any

comments, Debra?

Michael?

DR. SOMERS: You also have the TEP weighing in with the thumbs up?

DR. NARVA: Yes.

CO-CHAIR ANDERSON: Any further discussion on the part of the committee? All right, ready to vote on validity.

MS. OGUNGBEMI: Committee is voting on validity for Measure 1425. Options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are zero votes high, 19 votes moderate, zero low and zero insufficient. Measure 1425 passes on validity.

CO-CHAIR ANDERSON: Moving on to feasibility?

DR. NARVA: The data source is CROWNWeb, it seems to be quite feasible.

CO-CHAIR ANDERSON: All right, let's go for the vote.

MS. OGUNGBEMI: Committee is voting on

feasibility. Options are 1 high, 2 moderate, 3 low, 4 insufficient. This is for Measure 1425, voting is open.

Results are 15 votes high, four votes moderate, zero votes low and zero votes insufficient. Measure 1425 passes on feasibility.

CO-CHAIR ANDERSON: All right, moving on to use -- usability and use.

DR. NARVA: So this measure is not currently in use although it's an existing measure.

I actually would ask the developers or the potential users why it's not being used?

DR. ANDRESS: So this measure faces many of the same issues that we discussed for the pediatric -- the measurement of hemoglobin for pediatric patients.

I think it's -- you know, it's available. It's certainly there for use outside of CMS and in the meantime, we're trying to resolve how best to address quality measurement

in a critical yet very small patient population in the dialysis community.

So this is a measure that I think is still in the running for consideration in public reporting and other uses. We simply haven't managed to conclude the conversation at CMS.

DR. NARVA: The pediatric caucus want to weigh in on this?

CO-CHAIR CROOKS: So it's usable, but not in use in a sense? You know, it's usable, somebody could the data from you, the report or the results, but no one's working with you to use it yet, but it's usable?

DR. ANDRESS: I think that would summarize it correctly, yes.

CO-CHAIR CROOKS: Okay.

CO-CHAIR ANDERSON: All right, ready to vote for usability and use.

MS. OGUNGBEMI: The committee is now voting on usability and use for Measure 1425. The options are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is open.

Results are four votes high, 15 votes moderate, zero votes low and zero votes insufficient. Measure 1425 passes on usability and use.

CO-CHAIR ANDERSON: All right. Oh, one more time.

Any further discussion on the measure before we vote for suitability to recommend endorsement? All right, shall we go ahead and vote?

MS. OGUNGBEMI: The committee is now voting on Measure 1425's overall suitability for endorsement. Options are 1 yes, 2 no. Voting is open.

I still need two votes.

Results are 19 votes yes, zero votes no. Measure 1425 passes on its overall suitability for endorsement.

CO-CHAIR ANDERSON: I think we are at the conclusion. The last two measures we're going to have to take up next week as well as the related and competing measures that we didn't get

to. I'd like to thank everyone for the past two days, their time, their expertise and their engagement in the process. It's been a great two days. Thank you very much.

CO-CHAIR CROOKS: Thank you very much.

MS. BAL: Before everybody leaves though, there are some last minute timelines and such. So as a reminder, we're meeting next week on Tuesday 1:00 to 3:00. Please come ready to discuss the last two measures, 1460 and 1662.

Also, on behalf of the staff, we do want to thank the committee. It's been a pleasure working with you for the past two days and we really enjoyed how involved all of you are. Thank you.

So the meeting's department will be sending an email soon after this meeting to give you reimbursement forms and as Sarah said earlier, the hotel costs have been reimbursed already.

(Whereupon, the above-entitled matter went off the record at 2:47 p.m.)

B

C

D

F

M

Q

W

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Renal Standing Committee

Before: NQF

Date: 05-07-2015

Place: Washington, D.C.

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter