NATIONAL QUALITY FORUM

+ + + + +

SURGERY ENDORSEMENT MAINTENANCE 2010
STEERING COMMITTEE

+ + + + +

THURSDAY
MAY 5, 2011

+ + + + +

The Steering Committee convened in the John T. "Jack" Elliott Room at the Embassy Suites DC Convention Center at 900 Tenth Street, Northwest, at 9:00 a.m., Arden M. Morris and David F. Torchiana, Co-Chairs, presiding.

PRESENT:

ARDEN M. MORRIS, Co-Chair, MD, MPH, FACS
DAVID F. TORCHIANA, Co-Chair, MD
NASIM AFSAR-MANESH, MD
HOWARD BARNEBEY, MD

JAMES E. CARPENTER, MD
ROBERT R. CIMA, MD, MA, FACS, FASCRS
CURTIS D. COLLINS, PharmD, MS, BCPS AQ-ID
PETER W. DILLON, MD, MSc
RICHARD P. DUTTON, MD, MBA
PAULA R. GRALING, DNP, RN, CNS, CNOR

VIVIENNE HALPERN, MD, FACS
RUTH KLEINPELL, PhD, RN-CS, FAAN
JOHN MORTON, MD, MPH, FACS
TERRY ROGERS, MD
CHRISTOPHER SAIGAL, MD, MPH, FACS
ALLAN SIPERSTEIN, MD
RENAE STAFFORD, MD, MPH, FACS

CAROL WILHOIT, MD, MS
CHRISTINE S. ZAMBRICKI, CRNA, MS, FAAN

NQF STAFF PRESENT:

HEIDI BOSSLEY

JESSICA BOWER

HELEN BURSTIN, MD, MPH

ALEXIS FORMAN

MELINDA MURPHY, RN, MS

JESSICA WEBER

ALSO PRESENT:

JOHN BOTT *

DALE BRATZLER *

SHERYL DAVIES *

JEFFREY JACOBS

KATHY JENKINS *

WANDA JOHNSON *

FLORA LUM

PATRICK ROMANO

BARBARA RUDOLPH *

DAVID SHAPIRO

JEFFREY SILBER *

DONNA SLOSBURG

* Present via telephone

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

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

(9:00 a.m.)

CO-CHAIR MORRIS: Good morning, everybody. Thank you for attending our second day of the Surgery Endorsement Maintenance Steering Committee, and thanks for your work yesterday. Excuse me.

We'll just recap very briefly. I thought we had a lot of good ideas and a lot of good discussion, particularly yesterday morning, quite a bit of discussion. I think as the day went on we indicated that a little bit.

For today, what I'd like to ask of the Steering Committee members is that you continue to present the measures that you're assigned in a succinct way, and we definitely want to hear your opinion, so we want to hear your opinions about the criteria.

You've spent more time with these measures than any other person on the Committee, and we value what we have to say

about it. So, for the Steering Committee members who were -- who gave a little bit more information about how they viewed the criteria, that was very helpful.

So our meeting today is open, and is it time to start the phones? All right. Dr. Torchiana has a few words, as well.

CO-CHAIR TORCHIANA: I would also like to convey my welcome, and I think my reflection on yesterday would be that the discussion is where the richness of the face-to-face meeting is sort of brought out, so the discussion is why we're here.

I thought we had numerous excellent discussions yesterday. Some of them were hard to wrap up in a bow at the conclusion, but that doesn't negate the value of the conversation and the thoughts that came out.

I'd like to add to what Dr. Morris said that for the developers today we have a number of developers that are both the same as

yesterday and then some new ones, again, to try to be very succinct and pointed in your description of the measures so that you can present to the Committee why the NQF should endorse this measure and the value of the measure for the patient population, as opposed to really a more exhaustive history of the measure. That way we'll be able to more evenly address the agenda today, hopefully. So, thanks everyone once again.

CO-CHAIR MORRIS: I just want to add one other thing to that, and that is that for measure developers we'd like for you to only comment on your own measures. We don't really want to hear your comments about the other measures that are other discussion until the public and member comment period.

We're going to start today with two measures that were left over from yesterday, and we're going to do them in reverse order. We'll start with Measure 0301. I believe that CMS has already or CMS

representatives have already introduced this yesterday, so please take it away.

MEMBER ASFAR-MANESH: Good morning. So we'll start off by talking about Measure 0301, which is surgery patients with appropriate hair removal. This is part of the SCIP measures supported by the measures stored on CMS.

There are a number of exclusions just to keep in mind in the description of this measure. These exclusions include ages less than 18, length of stay greater than 120 days, laparoscopic surgeries, patients in clinical trials, and patients who perform their own hair removal.

As far as the importance to measure and report this, there is some evidence back and forth, just so that the group is aware, about the fact that appropriate hair removal may reduce incidence of surgical site infections. Actually, in the outpatient measure that's paired with this --

I'm sorry, not paired but a complementary measure to this.

There are some more citings about three randomized control trials as well as two systemic reviews that show this correlation. However, there doesn't seem to be absolutely clear evidence that this is important.

Having said that, the measure when it was first reported in 2005 had a rate of 91.5, and now for the last several quarters ending in Quarter 2 of 2010 is now reported at 99.6 percent, so very close to being topped off if not topped off.

So, really, I think the only thing for us to discuss and where I think we need the Committee's expertise is since the performance for this measure has been consistently high, should we continue to measure and report this with the thought that this is currently integrated into the work processes in the ORs, and so there is minimal performance gap and the flip side of that

being that if we stop measuring and reporting this, could we actually revert back to previous performance, which was sub-optimal.

This could also, putting it together with the discussions that we had yesterday, could potentially meet the requirements for an inactive measure, something that gets followed up. So I'll wrap it up there, and we'll kind of open it up to discussion as far as what the group feels as far as continuing to measure and report.

MEMBER SIPERSTEIN: I want to make a comment. You know, this is one of these SCIP measures, just like pre-op antibiotics and redosing, et cetera, and is often considered a bundle of activities to reduce surgical site infection, although there has been some evidence in the literature that implementation of the bundle in certain patient populations hasn't moved the needle.

So when these are reported, we tend to look at the individual metrics, but we

also tend to look at any given patient where the entire bundle has been properly implemented.

CO-CHAIR MORRIS: Are there any comments or questions?

MEMBER DILLON: I also think it's unlikely that we'll see any recidivism here, because certainly in our ORs what it did was it got rid of the razors, and I think everybody has now moved to either electric clippers or not doing it at all.

So I don't see a sudden resurgence of people going out and buying Gillette razors to try to get around this. I think it should be put to sleep or inactive or whatever we can do to it.

MS. MURPHY: You know, there was a comment within the work group when this was discussed, exactly to your point, that for economic reasons, razors being cheaper than clippers, that there could be a reversion.

CO-CHAIR MORRIS: We have -- we

purchase clipper bases, and then just the clipper heads are moved out, which are probably even cheaper than razors, but I don't think all the clipper bases are going to go in the trash. Every hospital I've been in has done that. I don't know if anybody has any different opinions.

MEMBER MORTON: I think, you know, even though it's up in the nineties, there's still constant push-back from folks who don't think that it works, because different studies have cropped up, you know, refuting the value of the individual measures. So, as much as it looks like it is topped out, I think this is something we still have to be vigilant about, and I wouldn't sunset it just yet.

MEMBER DILLON: But at the same time, there is growing information that none of these are truly affecting the incidence of wound infections, and so, again, why continue a, you know, a metric? Why continue a policy that really for all intents and purposes has

not been shown to be effective in what it was set out to do?

We have changed it. We have topped it out. I think we should be going on and focusing at more aggressive, you know, and better wound care metrics.

CO-CHAIR TORCHIANA: Is there an NQF bundle measure for the SCIP bundle?

MS. MURPHY: They're each endorsed individually, but one of the things that went through my head as you were talking about it being reported as a part of a bundle is whether or not there would be any potential for moving it to an inactive status if that's what the group wanted with retaining it as a part of a bundle.

DR. BURSTIN: So let me just give you a little bit of background here. This is an interesting issue. We're actually taking this issue of inactive endorsement status to our Board next week, so we'll know definitively shortly.

But there is, as Ray Gibbons, the Chair of our Cardiovascular Committee, liked to refer to them, the Hall of Fame measures, ones that perhaps have had their life, probably time to put them on a shelf, but the idea would be if there were any concerns about the -- that the rate might, in fact, drop, the idea would be is there a logical place to put some of these measures where they would be on inactive endorsement status, meaning people could do periodic surveillance to make sure the rates aren't dropping, but at the same time we're working through a process.

If the rates drop, we could bring it back through active endorsement status, but, again, I think the point that was just raised is really the key one. Is there clearly a relationship between this and the outcome?

We went through this fairly carefully this past year, Heidi will know, because we did this for a couple of the

surgical specialty societies, that the evidence was still strong with the exception of a couple of key areas, some GU areas and some neurosurgical procedures. Other than that, the evidence was still pretty sound.

MEMBER ASFAR-MANESH: Well, and just to clarify, this is -- this is still part of the CMS bundle, correct? So, I mean, as long as the SCIP measures continue to be collected -- it's going to be collected?

DR. BURSTIN: Yes and no, so CMS does retire measures, as well, and they view them as topped out, and I'm not sure if Dale Bratzler is on the phone this morning.

We've seen some indication that, at least for the value-based purchasing program for CMS, they are probably going to retire this measure. It has not yet been retired for the inpatient payment program, but that is an issue.

I do know that there -- it's on the list for IPPF-2 to be retired, so that

would be one. If CMS is going to retire it, it'll just go away. I do know Dale has mentioned there is a possibility that there is an all-or-none composite of all the SCIP measures. It has not been submitted to us yet.

So usually if NQF removes endorsement, within a period of time it's usually removed by CMS unless they -- in the new section they actually need to put something out in the Federal Register indicating why they want to continue to use an unendorsed measure.

MEMBER ASFAR-MANESH: And they did note -- I'm sorry. They did note that they might retire this measure but that the Joint Commission will actually be taking over, it sounds like, if CMS retires the measure is what's noted in the measure.

MEMBER MORTON: Could we make the suggestion that there be a bundle?

MEMBER CIMA: If you're going to

have a bundle, that means someone's going to have to collect the data, so if you're going to have the data, you might as well report it.

I mean, if you're going to retire the measure as a stand-alone measure, then leaving it in a bundle means you still have to collect the data, so why -- you know, either you say the measure's a good measure to stand alone and participate in a bundle. I mean, I don't -- it's one or the other.

MEMBER MORTON: I guess, as to Allan's point earlier, was there are some data to say that if you're compliant with all the measures, what it be most indicative is the culture of accountability and making sure you're doing the right things at the right time, rather than stressing just the one individual measure.

So I think there's a little bit of argument for maintaining it as a composite measure, along with the whole bundle, rather than as a stand-alone. There are data that

are supportive of it, and I think sometimes we're in the process of being whipsawed back and forth with the latest study that comes out, but I think the preponderance of the evidence shows that it does have some utility.

MEMBER WILHOIT: In the Work Group D materials there is a PDF for the composite measure that does include this. It's number 963. So we were sent it. I don't know if it's up or not, but it is in the package of materials.

MS. MURPHY: And it was withdrawn. It's not ready for prime time.

MEMBER ZAMBRICKI: I just wanted to mention, since I was looking at the antibiotic, which was part of this grouping and in response to Dr. Cima's comments, looking at some current literature there was a paper out of the JAMA within the last year looking at 400,000 patients.

It found that, actually, the global use of the measures did show a slight

reduction in surgical infection but very powerful statement about the individual item relationships are weak, lack clinical significance, and there was no meaningful association between adherence to the individual measures and decrease in post-op infection, so that would be a reason for the bundling.

I think it is a bigger issue for us to look at these, because they're all reported as individual measures, and it does seem like there is some recent -- there as another paper out of the VA really saying the same thing, that looking at five years of experience that there was not a correlation between these individual measures and decreased surgical site infection.

CO-CHAIR TORCHIANA: Just anecdotally, the issue with hair removal is more of a what not to do, rather than what to do, so it's eliminating shaving the day before that really was the big advance, and then

after that it's all fairly trivial.

CO-CHAIR MORRIS: Okay. So we heard some arguments for making this measure inactive, some arguments for bundling it, and some for keeping it. Is everybody ready to go ahead and vote, or any other comments? MEMBER ZAMBRICKI: Could you provide some guidance as far as voting?

MEMBER HALPERN: Exactly. So for voting, this would just say that it's, you know, endorsable by NQF, but if we wanted to say, "Yes, we think it's a good measure, but it should be retired at this time," how do we indicate that with the voting?

CO-CHAIR MORRIS: First, we would vote to -- correct me if I'm wrong, Melinda, but first we vote whether to endorse it or not, so whether to continue to endorse it, and then secondly we would determine whether we wanted to have it considered for inactive status. Is that what you're asking for?

MEMBER ZAMBRICKI: You had

presented, it sounded like, three options. One is to bundle, one is to accept, and one is to accept and make it inactive, and so I was just wondering, a yes-or-no vote.

MS. MURPHY: So the activity is to evaluate the measure based as submitted, and the way in which you did it yesterday and we talked about was around the issue of importance. Is it important to continue to measure this one in this case?

So you could indicate that it was not important based solely on the fact that it was topped out and then continue to evaluate on the other elements if we're looking at it in terms of potential for an inactive status. So importance would be the place where you would indicate whether you believe it is important to continue to measure.

CO-CHAIR MORRIS: Does that seem clear? Anybody still a little bit foggy on that?

DR. BURSTIN: One more point. We

can't evaluate a bundle, because it wasn't submitted to us. You could strongly recommend, regardless of what comes out from the vote, that CMS submit, and in this case probably an all-or-none composite of the key process measures.

CO-CHAIR MORRIS: Anybody -- so, Dr. Cima, you had some comments about why you thought that that might be less valuable to have an all-or-none bundled measure, compositive measure.

MEMBER CIMA: No, I think that the data strongly supports the fact that it should be bundled. I'm just saying, you know, I think it would be better just to not report it individually, but, I mean, one of the things if you're going to have a -- because people are still going to have to collect the data, so the question is should it be a stand-along reportable thing.

If you answer that question, it seems to be it's topped out. The question is

that this individual metric, does it need to be reported individually as submitted is one view, and I think it's fine as a bundle.

I think it's important as part of the bundle, but, you know, still I think is this question. You're going to be collecting the data. Do people still want to report it? It's a different issue by itself, individually. That's what my thing was.

CO-CHAIR MORRIS: Anybody else? No? Time to vote? I'm getting the signal that it's time to vote. All right, so if we think that it's topped out, if you think that it's topped out, that it's no longer important to measure individually, then you would vote no on this first one, and if you think that it should still be measured individually, you'd vote yes.

MEMBER ROGERS: I'm sorry, Arden. Isn't it true that if we think it should -- they're important and should be bundled, we would have to approve both of them

individually and then make a recommendation that they be bundled, no?

DR. BURSTIN: No, the policy that we have is that all measures within a composite need to be evaluated as to whether appropriate as a stand-alone measure or only as part of a composite.

So you could make the argument this one could be -- maybe has outlived its usefulness as an individual measure but would potentially still be useful now that you evaluated it only within a composite. I know that's confusing, I apologize.

CO-CHAIR MORRIS: All right. Let's go ahead and vote.

And I'd like everybody to go ahead and press your either 1 or 2 button again. Okay. So four for yes, 15 for no, and it's probably important to record at this point that that's -- that as a committee that we think that this is topped out as an individual measure, and that's the reason for the no

vote, and we'll continue through the rest of the vote.

Does the measure meet NQF criteria for -- are we going to go through the rest of the vote? Oh. Okay.

MS. MURPHY: Do we not need to go through the remainder -- if we're going to put it on inactive status, we need to have assessed that it still meets all of the criteria, so I --

DR. BURSTIN: This is relatively new territory for us, so I think that would be fine. The Cardiovascular Committee, they just figure they'd revisit everything they viewed as topped out before later and come back to it. If you're on a roll, go for it.

CO-CHAIR MORRIS: All right. So we're engaged in a work in progress here. How about scientific acceptability of measure properties?

And, again, please hit your button and then hit Send, aiming at Jessica. Then

one more time.

Ten say completely. Eight say partially. One says not at all.

Usability. Twelve say completely, five say partially, one says minimally, and one says not at all.

Feasibility. Thirteen say completely. Five say partially. One says minimally.

And then do we need some guidance here? I think we need a little guidance here. Does the measure meet all the NQF criteria for endorsement? Is this -- are we talking about the individual measure?

MS. MURPHY: Yes. We are talking about the individual measure here.

CO-CHAIR MORRIS: Okay, and we previously voted that it does not meet the importance criteria, so does that mean that we have essentially a no vote premise?

DR. BURSTIN: I think you're -- I don't know if there's a need to do this, but

I think one question might just be to get a read of the group, and we'd have to hand count or maybe just use this one. Just do a hand count of if the inactive status is an option, would you recommend it for inactive status, rather than --

CO-CHAIR MORRIS: I think what --

MEMBER HALPERN: My personal would be recommend this as an individual, as an inactive status but that we would recommend continued use in the bundle, that they would have to present us with a bundle.

CO-CHAIR MORRIS: All right. I think we all pretty much agree on that. Is there anybody that would like to make any countering comments? Okay.

MEMBER DILLON: I would just say that any component of a bundle still has to be supported by whoever the steward is. I mean, that's -- and that still begs the question of whether this should be in a bundle, and that has to be scientifically proven and presented

as part of a bundle presentation.

I don't think we should be saying we want this in a bundle. We should be saying it should be considered for a bundle.

CO-CHAIR MORRIS: Thank you for clarifying that. The next measure is 0515, ambulatory surgery patients with appropriate method of hair removal. Again, Dr. Asfar-Manesh.

MEMBER ASFAR-MANESH: Okay. So this is a similar measure to 0301 looking now at the ambulatory surgery patients, and this actually, just to clarify as a deferring point to the previous measure, looks just at all ambulatory surgery patients who had hair removal, so it actually excludes patients who did not have any hair removal with the measure still being ASC Quality Collaboration.

So as far as an importance to measure and report this, again, conceptually it's the same as 0301. The measure stewards actually went to some great lengths to make

sure that it was harmonized with 0301 and added a number of different components, which they detailed to make sure that that happens.

The importance to measure is that we do have, as you said yesterday, 80 percent of surgeries that are now happening in the outpatient settings, so if we are doing something in the inpatient setting it makes sense that it actually gets carried to the outpatient setting.

They presented some data analysis that was done. So to give you an idea of the performance on this measure currently, they looked at 192 ambulatory surgery centers between July and September of 2011.

The performance on this measure was somewhere between zero to 100 percent, but the mean was actually 96 percent, and the median was 100 percent, so again a pretty high level of performance. There were basically 7.3 percent of centers that presented data that was less than 100 percent, so that would

really be the performance gap that we would be looking at.

So, again, as far as points of discussion, I think very similar to some of the things that we discussed in the inpatient measure as far as being close to being topped off. I was a little bit unclear about the validity testing that they did, just to direct your attention to that on 2c.

The validity testing was done by having a questionnaire go out to six nurses asking them five different questions, and this is not my area of strength, so I would appreciate some input from those in the group if you feel that that was appropriate and enough. Otherwise, I'll leave it open to discussion.

They do not -- another point is they do not have disparity data, and they explained that as if this is something that's federally reported, then they would have access to that and could provide us with that,

but that's another area that's lacking in this measure.

CO-CHAIR MORRIS: Any other comments, questions?

MEMBER DILLON: I would say I'm not sure that I want to contribute to surgical confusion out there if we have just inactivated one. Then you walk into your surgery center, and now they're counting, you know, how you shave.

So I would just point out that if we go one way on one and one the other, we're going to create, you know, surgical confusion, and surgeons don't need to be any more confused than they are on some days.

MEMBER STAFFORD: I think you're right about that. The only caveat with that is that there are a lot of freestanding ambulatory surgical centers where the surgeons only operate there, and they don't operate in a hospital. So, while you're right for those of us who work in both worlds, it would be

confusing to have them both.

There are places where they don't, and that would be the data that I would be really interested in from the developers is do they know when they look at their data in terms of disparities how many of them are freestanding centers and how many are associated with academic centers, because I bet there is some difference in the data.

MEMBER HALPERN: I also don't understand why they don't occlude people who didn't have hair removal.

MEMBER ASFAR-MANESH: So they actually explained that. There's about 75 percent of outpatient surgeries are actually ones that would not require hair removal, so the two big categories are cataract surgeries and injections for pain. That's why they actually -- they just wanted to decrease the burden to people who would be collecting it by taking out the 75 percent of outpatient cases.

MEMBER HALPERN: So why don't they

take out those ones that don't require hair removal, rather than taking out all those who didn't have hair removed, because then you might be missing somebody who didn't have their hair removed?

CO-CHAIR MORRIS: Can the measure developers speak to that?

MS. SLOSBURG: I can speak to a couple things. One is the first issue regarding topped out, and I think we talked about this yesterday that ambulatory surgery centers do not have any federal mandate to report.

So in our database right now we're up to about 800 ASCs, which is less than 15 percent, and, as you were saying, there are a lot of surgery centers the physicians just go to the surgery center and not to the hospital. So I think if you did take a look at the other surgery centers, I don't think the compliance rate would be as high.

When we started out, we were

similar to CMS where we had a low rate, and each quarter we improved. So that speaks to topped out, so we would like to have this measure so that when we do have a federal mandate we can look at all surgery centers and see where we are, and if we are, then we can move on.

Regarding the question about the no hair removal, you know, we just never, I guess, never thought of it that way, but we certainly could look at it that way.

We did try to harmonize with the hospital, and the main reason was because of cataract and GI being such a large portion. That would mean that someone would have to actually go and count those patients, and that's a huge burden on ASCs.

It's a lot easier to look at who had hair removal and then walk through those medical records than having to pull all those cataracts and all those GI cases when you know there was no hair removal, so that's the

reason for it. We also uses ICD-9 versus CPT. Sorry.

MEMBER HALPERN: Why not just eliminate those cases that you know don't have hair removal and look at everything else, though?

MS. SLOSBURG: I mean, my understanding would be if it's in the numerator or the denominator, then you have to actually look at those cases to say it's a no, correct?

MEMBER HALPERN: I'm saying exclude those cases that don't require hair removal. Make that your exclusion instead of making no hair removal your exclusion.

MS. SLOSBURG: Okay. I mean, we can certainly look at that.

MEMBER CIMA: I mean, I sort of -- I mean, I certainly agree with Peter on this, but the issue becomes this is directed at surgical site infection reduction, okay. There is no data in there about ASC surgical

site infection.

MS. SLOSBURG: There is no data out there.

MEMBER CIMA: I know. I know. I'm saying that as just a point of fact. There is no data available, so you have no idea if this is going to have any impact. We have plenty of data to say individual measures have been reported in the SCIP criteria, have not been shown to be effective.

So why -- and we just voted to inactivate because of lack of importance in the inpatient setting. So this seems to be a measure just to have a measure that follows along a measure that we didn't think was an adequate measure, and so should we put ourselves through that?

MS. SLOSBURG: Well, I --

MEMBER CIMA: What's the utility of this data if as itself it doesn't fit and you have no data to show that surgical site infections is even a problem in the outpatient

setting? So how are you going to show improvement?

MS. SLOSBURG: Well, the only -- I'm going to let Dr. Shapiro speak to that. The other thing is that in the SCIP measures the only other measure that would really be appropriate to ASCs is the IV antibiotic timing, and we did try to harmonize with that measure. We could bundle those two.

MR. SHAPIRO: I think the only -- I think the reason that we would really request that you continue to endorse this measure in the ASC setting, even given what I've -- I heard that very good discussion about the one before, the inpatient and outpatient settings -- is where this industry is in reporting and doing their data collection.

All of the data that you see is only based on less than a fourth of the surgery centers that are Medicare-certified out there. What we've tried to do is

harmonize measures with the existing CMS measures.

I realize today may be the beginning of a sea change, but I would urge this Committee to continue endorsement of this measure to allow us to get to come back to you with some better statistics on a greater proportion of ASCs, because my fear as a clinician is that there is a lack of penetration in the ASCs to the extent that there is in the larger surgical community of these techniques and of the importance of these techniques in preventing HAIs and surgical site infections.

So, for those reasons, I think because we are at a different place than the other facilities about which you changed your endorsement prior to this, I really urge this Committee to let the ASC community continue to use this as a measure to assess our ability to conform with what at least previously has been shown to be process measures that will reduce

infection in our patients.

MEMBER DUTTON: From the national anesthesia registry, two-thirds of all surgery centers are hospital-based and probably fall under both sets of guidelines. I would strongly recommend keeping our thinking about this in line with our thinking about the previous measure to reduce confusion. The ASCs can have the same bundle that the inpatient facilities have and arguably should have the same bundle.

CO-CHAIR MORRIS: I think another point that you made yesterday, Richard, and that I sort of expected for you to make today was that endorsing these measures is not a means to collect data for research to see where we stand.

It's supposed to be based on actual data that indicates that we'll be able to make a difference in quality. So I think that you make a strong argument, and clearly you have the best of intentions, but the goals

of NQF endorsement are a little bit different than that.

MEMBER MORTON: Well, I agree with what Peter and others have said that if we didn't vote the other measure, we shouldn't vote this one. It would be inconsistent.

MEMBER WILHOIT: One question I had is about some of the differences between this measure and the other one, and that was that in the denominator those patients who perform their own hair removal are excluded, but it seems like if a razor increases the risk of infection, it would increase the risk regardless of who does the shaving.

And then in the numerator, patients with hair removal from the scrotum are handled differently, and they weren't in the other measure, and so I wasn't sure why those differences were there.

MEMBER ASFAR-MANESH: So the scrotal surgery actually is the same in the inpatient one, as well. The inpatient one

actually has an exclusion for neurosurgery and scrotal, and, actually, the measure developers try to harmonize with the inpatient measure to add the scrotal excluded through neurosurgery, because they don't have neurosurgery cases in the outpatient setting.

MEMBER WILHOIT: Although it's handled, instead of being handled as an exclusion, it's handled as a numerator-positive, so it's handled in a different way, rather than lining up.

CO-CHAIR MORRIS: All right. Let's go ahead and vote. Does the measure meet NQF criteria for importance to measure and report, and, as before, if you think that it's topped out or that it should be part of a bundle but not stand-alone, then you would vote no. And if you believe that it should be a stand-alone measure, then you would vote yes. We have six yes votes and 13 no votes.

Because we've talked about it as a bundle, I think that we should continue with

this vote, since we're sort of inventing how this is done. So, next, scientific acceptability of measure properties. Five say completely. Thirteen say partially. One says not at all.

Usability. Seven say completely. Nine says partially. Two say minimally, and one says not at all.

Feasibility. Thirteen say completely. Four say partially. Two say minimally.

And then here again, as before, I think that we don't actually need to vote on does the measure meet all the NQF criteria for endorsement, because we voted no for importance. So I think that the Committee would like to respectfully recommend that this be considered for a bundle measure by the developers.

Okay, and so our next step is a brief introduction of the measures for today by the developers, and some of the developers

from yesterday spoke about measures for today, but many developers are getting their first opportunity to speak, so we'd like to go ahead and start with the developers for the American Academy of Ophthalmology, et cetera.

MS. LUM: Hi, good morning. I want to first thank you for the opportunity to speak, and I want to give credit, actually, to the NQF committees for stimulating our interest. They really strongly urged that we look at patient-reported outcome. We already have two cataract clinical outcome measures, so these are amplifying it.

Do you -- I'm sorry, a question. Do you want me to address both measures at once? I don't know if that would be a little bit briefer, or else do you want me to do it separately?

CO-CHAIR MORRIS: We'd like you to develop -- to address both at once and to be pretty succinct with it.

MS. LUM: Okay. Great. The

visual function measure we believe fits into the National Priority Partnership's priority of population health by providing an index of visual health. Patient satisfaction fits perfectly with the patient and family engagement.

We also think that these line up really well with the NQF goals. Visual function is an AHRQ-tested and validated instrument that provides additional information on the role of visual impairment other than visual acuity, as we said, which is a clinical measure, and that's measured by the clinician.

This is a patient-reported measure. How do patients do on a daily basis reading small print, reading a book, seeing steps and curbs, reading traffic and street signs?

The patient satisfaction measure could be added to the NQF portfolio of patient experience measure, which already includes the

CAPs, the HCAPs, and the family evaluation of hospice care. This was approved by the CAPS consortium and developed with the same scientific rigor and standardization as all the other CAPS instruments.

The survey asked patients about information that they received prior to surgery, including risk-benefits, alternatives, information about their care during surgery, post-surgical instructions, and the behavior of the surgeon -- Did they listen? Did they spend time with the patient? Did they allow them to ask questions? -- and an overall rating of the quality of the surgeon. So, in terms of importance, as you know, cataract surgery, the most performed procedure in Medicare beneficiaries, about three million beneficiaries, but there is no systematic approach.

So we think that these two measures really complement the clinical

measures that we already have, post-op complications, which we consider a never event, kind of a measure of surgeon proficiency, the visual acuity measure that we have already that really talks to appropriateness of care, the patient population that really should be operated upon.

Then now we have a visual function initiative that talks about what really matters to the patient, what affects them in everyday life, and then patient satisfaction as a measure of the patient experience with care, and we think that surgeons can really look at the results of these measures in a comprehensive picture of what happens in these patients and enhance patient outcome and their interaction with patients.

I did want to also address some specific issues that came up in the work group discussions last week. The visual function measure, there was a comment that it was a

complex numerator, and that's because we approached it, I think, thinking of it more as a CMS measure for PQRS.

CO-CHAIR MORRIS: I'm going to -- I'm actually going to ask you to stop there --

MS. LUM: Okay.

CO-CHAIR MORRIS: Since we just wanted you to introduce the measure. You'll have an opportunity to speak about these things in just a few moments. Dr. Barnebey, would you like to go ahead and present?

MEMBER BARNEBEY: Okay. Good morning. So there are two measures. You want me to address them together or separately?

CO-CHAIR MORRIS: You can address them separately.

MEMBER BARNEBEY: Okay. So the first one looks at using a survey, a VF-14 or, actually, a shorter version of it, which has actually been validated scientifically. It's a pretty strong measure looking at patient outcomes, so this is different than the other

measures, which is more of an objective measure.

So in terms of importance, you know, cataract surgery is perhaps the most frequent surgery done in the U.S., at least for CMS purposes, and they're looking at in terms of, I guess, looking at the gap measure maybe of about 90 percent patient satisfaction, so there is still an opportunity to improve things further.

The scientific data, you know, looks strong. I didn't see anything that was prospective studies, but, you know, there was some -- the research there and the scientific validation of the instrument looks strong.

One of the questions is usability, and I think the way the information was presented, if it's segmented out in terms of groups that have no comorbidities and those which do have comorbidities, I think that will be more useful.

Then, finally, the issue is

feasibility, because the data there suggests that these particular surveys were administered, you know, by a trained person, whereas doing the surveys by mail or electronic, I didn't see the science out there to show that that's there, so there's some potential in terms of a feasibility issue, but overall I thought it was pretty strong.

MEMBER STAFFORD: Excuse me. Are you talking about 1549?

MEMBER BARNEBEY: No, I'm talking about 1536.

MEMBER STAFFORD: So that's a survey?

MEMBER BARNEBEY: Yes.

MEMBER STAFFORD: Okay. Thanks.

CO-CHAIR MORRIS: Any other comments or questions that came up in the work group meeting from the Committee members?

MS. MURPHY: There was in the work group meeting some discussion about the fact that it is a self-administered survey that's

been validated for assisted administration, and there is a notation in 4e1 in the documentation that the elderly would potentially need assistance in order to be able to complete the survey. It also, in terms of burden, speaks to the potential necessity of a need for follow-up with individuals in order to get them completed.

MEMBER MORTON: I guess the only question I had is if we do the survey, what's the actionable item about the survey? What do we learn? What do we take back to improve our practice and care of patients?

MEMBER HALPERN: And I would ask, if it's the patients administering their own survey, is there a validation with true visual acuity as measured objectively?

MEMBER BARNEBEY: And, I guess, when I was reading this, I was looking at this as maybe a paired measure with the other clinical outcome, which was 0565, so there is already a measure out there looking at the

visual acuity as an outcome, as well as complications of surgery, and I would look at this, if we look at harmonization, as a paired measure, you know, not to replace it.

CO-CHAIR MORRIS: Can the developers tell us whether or not they intended for this to be paired?

MS. LUM: Well, actually, originally we had submitted a composite measure, which would have been the two clinical outcome measures, as well as these two patient-reported outcomes, but because these measures had not come up before as individual measures, we were advised to submit them as individual measures, but the intention has always been from the Academy's side as a composite.

CO-CHAIR MORRIS: Okay.

DR. BURSTIN: How do you define improvement? Like what's the degree on the scale of improvement is a question I had asked earlier.

MS. LUM: The VF-14 is scored from zero to 100, so 100 being perfect that you've done -- that you can perform all the activities, zero being not. So the improvement would be a score -- an improvement from the score pre and post on that scale.

DR. BURSTIN: If you went up two points, that would count as an -- I mean, I'm just trying to get a sense of the scale.

MS. LUM: There has not been a defined step improvement. I think we could stipulate it as part of this measure. As part of the scale and how it's been used throughout ophthalmology, there has not been a defined, I guess, interval that would be considered improvement.

The other thing I was going to say is there have been two studies that have looked at self-administered VF-14 compared to interview-administered and found that that was also a valid way to administer the test.

MEMBER BARNEBEY: If I could ask a

question of the developers, my understanding is you're suggesting not the VF-14 but the VF-8r, instead, which is a shorter version, but it's been clinically validated to be acceptable.

MS. LUM: That's right. For that feasibility issue, patient response, patient ability to return the questionnaires and response rate, we have recommended a shorter version. As you know, there is a rich literature on visual function and several instruments out there, but we were advised also to propose one instrument that would be used for this measure.

CO-CHAIR MORRIS: Can you address the other comments and questions that the Committee has raised so far before we ask you some additional questions?

MS. LUM: Sure. I think your question is about how can this affect clinician behavior, and I think the whole rationale of cataract surgery is to improve

visual function.

So if you find the patients without noticeable or significant improvement on visual function, then that will stimulate the surgeons to look back on those patients and see what characteristics about them would have maybe prevented, should have prevented them from being selected for cataract surgery or if there were any other factors that might have been able to be ameliorated prior to surgery that would have improved their outcome after surgery.

CO-CHAIR MORRIS: I think there was a question about -- I'm sorry. We will, I promise, we'll get to you, Dr. Rogers, but I think Dr. Morton had asked a question about actually burden, didn't you? Wasn't it about burden on the provider?

MEMBER MORTON: Burden on the provider, and I guess after hearing the comment I'm just wondering what's the incremental gain of the survey in addition to

just the clinic visit, where I would think you would get some idea of function at that point?

MS. LUM: I guess the advantage would be that it's a standardized instrument, and it asks about several realms of visual function, not just one. The clinician may not have time to ask every patient in a systematic approach, and the other question was about burden, and that's why I think we've addressed the shorter questionnaire.

People have done this in practice. I mean, it has been mainly used in a research setting, but I think because it's so important, a patient's improvement after surgery, I don't think it poses an undue burden on the provider or the patient.

MEMBER DUTTON: I think, in some way, this is a way of quantifying that conversation between the doctor and the patient. I think we're always going to have trouble where science collides with patient-centeredness in these measures, but I think

the patient-centered ones are very important. I mean, if you go have your cataract fixed, and I have, you want to see better afterwards.

CO-CHAIR MORRIS: I am sorry. Dr. Rogers, would you like to have --

MEMBER ROGERS: Just using the Wilhoit approach here, 2C1, there's probably just a typo that addresses the return rate on your validity testing. I'm concerned that of 414 patients, only 210 returned the questionnaire.

It is good for a survey, but it may not be good for the business that we're in, and I have concerns about that. You know, normally, people who are pleased will return or have some specific complaint, so we have no idea about the other half.

Secondly, there's an issue about 51 returned the VF-15 post-operatively. I assume it's a typo, but I just want to make sure there's no other secret method you have out there that's named the 15, as opposed to

the 14.

MS. LUM: No, I'm sorry, that must be an error, but, yes, the return rate I think has been looked at 50 percent, which is high for a questionnaire, but we're hoping if it's for measurement purposes the physician's office would probably undertake a more directed effort to try to get those patients to return their questionnaires. They'll see them back at post-op visits, and they can kind of ask them or badger them for their questionnaires.

MEMBER CIMA: I mean, that's a big issue in the exclusion criteria for the denominator, using Carol's approach. It says, "Patient refuses to participate." How do you define -- I mean, if they just don't return it, then they exclude it, so then you're cherry-picking people for your survey results.

The other thing is that, getting to Richard's point, though, is if you have an objective measure of visual performance, then

do you -- I mean, you can ask the patient are they seeing better, but you also have a measurement of them seeing better.

I mean, they go hand-in-hand, but one that's a flawed measure to be paired with one that's an accurate measure, I'm not sure it's -- does it add value?

MS. LUM: Well, actually, some people would say visual acuity is not a complete measure of visual function. Visual function encompasses a lot of things that aren't included in visual acuity such as peripheral vision, visual processing, contrast sensitivity, glare, acuity under glare conditions.

So visual acuity actually is a very -- it's just one dimension of visual function, whereas this takes into account a lot of different dimensions of visual function, actually gets to what does the patient do every day that's affected by vision.

MEMBER DUTTON: I'd agree with that. You can turn the argument exactly around and say that the strict science of measuring, you know, focus in the eye, isn't really capturing what the patient wants out of the surgery.

MEMBER ROGERS: And if I can make a comment, I don't think there's anything to suggest that the VF-14 as an instrument is flawed. There really isn't any science to support that.

I do have a question to developers. There are new implants that are now being used in a small group of people, multi-focal implants, and I wonder if 90 days is an adequate period of time to really be measuring that group of people. I was wondering if that was considered or discussed when you developed the measure.

MS. LUM: You're right, because, obviously, that wasn't considered in the development of the VF-14, and I think the

questions are valid, but you're right. The time period, I think, to harmonize with the other measures, have been 90 days. Ninety days is the visual acuity measure, so we would have to think, I guess, about how to keep that harmonization if we change it from 90 days.

MEMBER ROGERS: Because that could really skew your results.

MEMBER WILHOIT: I like this measure, and I like that it's survey and that it's patient-centered and so on, but it just doesn't seem very well defined.

First of all, some of the comments were, "Well, the physician could badger the patient to get the survey in." Well, that right there adds bias. The method by which the survey is performed is really key.

So, for example, for assessment of health plans, when the CAP survey is done, we have to hire a vendor, and we can't even know who the patients are who were surveyed. The vendor does the sampling, and we can't even

know who those people are so that we're not badgering the person to get the surveys in. So I think that that becomes really key, because if I badger my patients and you don't, we may have different -- we may have bias based on that.

Second, I think, you know, as specified, and this is what Helen raised: what's improvement? Well, the numerator is based on improvement, and is improvement one point or five points or ten points? Until that's defined, we don't even know what we're measuring.

Also, if you look at the numerator, and that's in 2a3, there's, "The numerator includes," is what it says, although it seems like some of these would be exclusions, "patients who did not complete their visual function assessment within 90 days."

So if somebody didn't complete the assessment, you get credit is what it says, so

you get credit for improvement even if you don't know, and then D is patients who did not have an improvement in their visual function, and there is no documented medical or patient reason for not doing so.

Well, there are so many nots in that, I'm not sure what it's saying, but it doesn't seem like it would be a numerator event. And then, in the denominator, it includes all patients, but for a survey measure there is burden.

There is cost to doing a survey, and particularly if you require a vendor or something like that to avoid bias, so consideration of sampling might be a possibility for high-volume centers, but that's something to think about.

The numerator time frame is a year, and yet it's specified as being 90 days, so there is a disconnect there, and then under exclusions, 2a9, is there -- it's actually 2a10. If you document a patient reason for

not improving visual function or document a medical reason for not improving visual function, the person is excluded.

Well, that sounds to me like if there's a complication, I can code that there was a reason, which was a complication, and suddenly the person is excluded. So, I really like this measure, but I think it needs a lot more refinement before we're ready to assess it.

CO-CHAIR MORRIS: Okay. So just to recap the Committee comments, first of all, it's unclear what really constitutes improvement in this scale. Secondly, the burden on provider is unclear.

The developer said that they didn't think there would be too much burden on the provider, but we know that there is a time and effort and also a cost component. Also, there are concerns that the provider going after the patient to complete the survey may bias their survey results.

There was some lack of clarity in the exclusions, including maybe too extensive exclusions, and then, lastly, the numerator time frame was disparate within the language of the measure.

Any other comments before we vote? And, you know what? I should also say the positive side. Sometimes I leave that out. So there is enthusiasm for the patient-reported side of this, patient-centeredness, which is certainly a laudable goal and one that we need to continue to push forward on.

So it's not that the measure in principle is not regarded favorably. It is regarded favorably, but just there are some problems within it that could use fixing.

MS. LUM: Thanks. I just wanted to comment. In terms of the numerator, I'm sorry if it wasn't clear, but, yes, the whole reporting period would be over a year, but it would be 90 days that the patient would be -- by 90 days, the patient would be asked for

their visual function or satisfaction reason.

The other part of it is, as I said, the numerator was complex, because we approached it as a reporting measure. If it was a performance measure, yes, we would only count those in the numerator who had improvement in visual function or satisfaction.

And we included the others because, I'm sorry, that's -- we had been in that mind set because of how PQRS measures are constructed and that when you report, you get credit for everybody, even if they didn't actually show an improvement.

So, sorry about that. We could simplify, definitely simplify the numerator just for the reporting purposes.

In terms of the exclusion, we thought it made sense that if the patient refused to participate, I think that would be at the outset, not because they turned out to have a complication after cataract surgery or

that there was a medical reason to do the cataract surgery, not because you want to improve their visual function after cataract surgery but because there is another condition or just to visualize the back of the eye.

We know that the vast majority of cataract surgery doesn't include those exclusions, so, I mean, I think we could support a case that there would be no exclusions in this measure that would simplify it.

We know there's a few cases, but overall we think the vast majority of surgeries we could look at the measure that way, not expecting a perfect 100 percent but knowing that those exclusions could go away.

CO-CHAIR MORRIS: Thank you for clarifying that. Now we're going to go ahead and vote on the measure as written. Does the measure meet NQF criteria for importance to measure and report? Eighteen say yes. One says no.

Scientific acceptability of measure properties. Two say completely. Twelve say partially. Four say minimally. One says not at all.

Usability. One says completely. Fifteen say partially. One says minimally. Two say not at all.

Feasibility. One says completely. Twelve say partially. Fourteen say minimally. Two say not at all.

Lastly, does the measure meet all the NQF criteria for endorsement? Nine say yes, and ten say no.

DR. BURSTIN: Especially since this is one of the first examples of a patient-reported outcome like that, so I think it would be really useful if the Committee had specific suggestions or questions back to ophthalmology that they could work on.

I mean, for example, we do have a similar measure on the depression side now of the use of the PHQ-9, which is the classic

scale we use for depression at baselines. There's a process measure that says, "Did you do the PHQ-9?" Then there is a measure that says the actual rate at zero, six, and 12 months.

So I think there are some interesting areas that I would hope the Committee could give advice, because this is a great direction we really want to go, and I think it just really comes down to the scientific acceptability of measure properties. Is it really ready for prime time?

MEMBER WILHOIT: I would say cleaning up the numerator, cleaning up the denominator, cleaning up the exclusions I think would -- I mean, I'd love to look at this again on a follow-up call. I voted no, and I was really close, and if it were a clean measure, I think we'd want to go -- I'd want to go for it.

CO-CHAIR MORRIS: So, we can

revisit this on a follow-up call if you guys would like to work with it some more.

MS. LUM: Definitely. Thank you.

CO-CHAIR MORRIS: Okay, and then the next measure, also, Dr. Barnebey.

MEMBER DILLON: Just on that last one, don't forget to add in what a meaningful scale is in terms of response.

MEMBER CIMA: And the methodology.

DR. BURSTIN: Exactly. It's still not clear, for example. Is it a delta measure? Is there a pre-op number, a post-op number? Is there a delta we're looking at? It's just it's still -- the numerator is still very, very fuzzy.

MEMBER CIMA: Also the survey methodology. I mean, that's a big issue here. I mean, I don't know if the developers really have contacted their membership and see the burden that an individual practitioner is going to have to go through.

I mean, most of these are done at

an ASC. You're not looking at, you know, a big hospital system that has, you know, other things they're doing.

You're looking at a guy or girl out on their own. They've got one person doing their scheduling. Now you're going to ask them to do this. I mean, it could be a big burden to do it appropriately.

CO-CHAIR MORRIS: All of the comments from today will be transcribed and posted, as well, so you'll be able to refer back to them.

MEMBER BARNEBEY: Okay, the second cataract measure being proposed is 1549, looking at patient satisfaction following cataract surgery. In terms of, you know, importance, obviously we talked about cataract surgery being a very common occurrence and something that needs to be measured and perhaps improved upon.

The one thing that this measure looks at which is different is, again, looking

at more the patient's perspective on their experience. As opposed to looking at visual function, this one is more of an experiential type of measure, and I think, in terms of the scientific merit, it was harder for me to put my arms around this.

There is a model out there that apparently has been clinically validated, but I wasn't familiar with it, and I was wondering, of the different types of surgical especially referenced, you know, how many of those were particularly applicable to ophthalmology?

In terms of useable, I had problems with the math in the equation with this one, as well, so I guess I need some help in terms of how that would translate into a useable measure.

Then feasibility, again, it's the burden of getting the information and that sort of thing, so this one I was a little less clear on. I clearly see the value of it, but

it's a new process for me to look at and I think for most of us to look at, as well, so I have some questions specifically for the developers.

The two particular questions are the modeling that you did in terms of the questionnaire that was field-tested that you worked in conjunction with the American College of Surgeons and also the mathematical model that you developed. To me it seemed a little confusing.

DR. BURSTIN: And I'll just indicate that we did go to AHRQ to ask them to actually submit the surgical CAPS tool they've been working closely with ACS on that. They don't feel ready to do that at this time.

CO-CHAIR MORRIS: Any other questions or comments for the developer? Okay.

MS. LUM: Yes, in terms of the CAPS, the surgical CAP, we did it as a cross-surgical collaboration with the Surgical

Quality Alliance, so you're right. It's not specific to ophthalmology, but we feel that it can address a lot of the important concerns about did the surgeon listen to the patient, did they -- was the patient asking questions, were they provided the pre-op instructions, post-operative instructions.

In terms of -- I guess I'm not sure in terms of the mathematical model. What did you -- were you referring to?

MEMBER BARNEBEY: In terms of understanding how the results were presented, it just wasn't to me intuitive when I was looking at the measures and how they were broken down and presented.

MS. LUM: So this is how the CAPS is -- SCAPS is scored, then, so there's different composites for how the SCAPS is scored. I have them here.

Then you would look across, so it's information to help you prepare for surgery, how well surgeons communicate with

patients before surgery, surgeon's attentiveness on day of surgery, information to help you recover, how well surgeon communicates with patients after surgery, helpful, courteous, and respectful staff at surgeon's office, and an overall rating of the surgeon, which is zero to ten.

Then there's different ways of presenting the data, but basically the proportional scoring method is there's three options for the responses, yes, definitely, yes, somewhat, and no, and you would just calculate the average proportion across the category for each composite. That's how usually -- that's how the SCAPS would be -- the results of the SCAPS would be shown or displayed.

MEMBER WILHOIT: I think this may be another one where the numerator statement isn't at all clear, because the numerator statement doesn't reflect any of that. The numerator statement also, again, seems to

include -- it seems like the rate would almost come out to be 100 percent. So the measure is satisfaction, but what seems to be measured is whether you measured satisfaction.

Well, those are very different. This is being presented as a public accountability measure, and whether you assessed satisfaction is of no interest from a public accountability standpoint.

So, again, I think the concept is one of great interest, but I'm not at all clear on exactly, you know, what is satisfaction. Is that -- you know, what is -- what counts as satisfaction? Is it one question? Is it ten questions? Is it a composite? Is it a -- you know, what's the score? I think there's just lots of questions.

CO-CHAIR MORRIS: Any other comments or questions for the developer? Would you like to say anything about what counts as satisfaction? Have you all

discussed that in your development group?

MS. LUM: Yes, and I'm sorry for the confusion again in the numerator. It definitely is what is satisfaction, not that you just measure satisfaction, but it would be the result that patients were satisfied after cataract surgery.

In terms of the SCAPS, again, it's like the VF-14. They haven't defined, I guess, a base level of satisfaction. Definitely we look at the proportional scores.

You'd want the yes, definitely, yes, somewhat, obviously, the majority versus the no responses in each of the composite measures and then the rating of the surgeon, which is zero to ten, ten being the best. Obviously, we want it greater than five, but maybe we can also regroup and try to clarify those things in the measure after this meeting.

MS. MURPHY: Would you mind to say just another word of when you say composite

measures, to what are you referring specifically?

MS. LUM: So the SCAPS consist of 41 questions, so they're broken out into, you know, pre-surgical, during surgery, after surgery, and staff, so those are the composites -- sorry -- that I was referring to that SCAPS breaks it down into.

CO-CHAIR TORCHIANA: I had a question on the feasibility issue. How universal is membership in the American Academy of Ophthalmology? Is that a very general group that virtually all ophthalmologists belong to?

MS. LUM: Right. We have 94 percent of all practicing ophthalmologists in the United States.

CO-CHAIR TORCHIANA: Could I make a suggestion that maybe the Academy could work out a way of becoming the vendor for the survey and that that might facilitate it for the broad practice?

MS. LUM: That is a good suggestion. We do have a PQRS registry, and my thoughts also were to serve as a vendor to serve as the web administrator of the surveys and also be able to score them and aggregate the scores.

MEMBER WILHOIT: Again, I think it's worth considering whether really it's all patients who had surgery or whether it's a sample. Sampling certainly helps control cost.

MS. LUM: Similar to the visual function, which I guess we had envisioned these as composite measures, and under the PQRS it is a sample of patients, just 30 patients, as well, so I think the burden requirements, we had been thinking that it would be a sample of patients and not all the patients.

MEMBER WILHOIT: But the denominator statement is all.

CO-CHAIR MORRIS: Okay. So let's

move on to a vote of the measure as written. First of all, importance to measure and report. Thirteen say yes. Six say no.

Scientific acceptability of measure properties. One says completely. Ten say partially. Five say minimally. Three say not at all.

Please hit your buttons one more time and hit the Send button. Three say completely. Ten say partially. Five say minimally. One says not at all.

Feasibility. One says completely. Ten say partially. Six say minimally. Two says not -- two say not at all.

Does the measure meet all of the NQF criteria for endorsement? Five says yes, and 14 say no, and I think that this is very similar to the previous measure that we would be interested in seeing a revised version of this measure and that we think that patient-reported outcomes are very important.

We're going to take a short break,

15 minutes. Let's reconvene just before 10:30.

(Whereupon, the foregoing matter went off the record at 10:14 a.m. and resumed at 10:33 a.m.)

CO-CHAIR MORRIS: We're going to go ahead and get started again, and we're moving on to the General, Prophylaxis and Wound Dehiscence section. The first measure that we'll be discussing is 528, and that will be discussed by Dr. Collins.

MEMBER COLLINS: Yes, okay, thank you. So a subject matter pretty near and dear to my heart here. The first is 0528, which I believe is a SIT measure number two up for re-endorsement.

This particular measurement is the surgery patients who receive the correct prophylactic antibiotics consistent with current guidelines based on their procedure, and the procedures listed in our packet are CT surg, vascular surgery, colon surgery, hip and

knee, arthroplasty, and vaginal and abdominal hysterectomies are listed.

There's a -- there is a comment about an appendix. However, I didn't see that for other procedures, but I know that other procedures such as neuro surg and other procedures are being rolled out as we go.

I think this is a very important measure. Selection of the appropriate antibiotic I feel goes without saying, as well as the importance of not selecting agents that are too broad from a collateral damage standpoint.

So, not only is there good evidence for based on the antimicrobial susceptibilities but also the spectrums of the agents, so I do think it's important, and the work group did, as well. The science behind it I think is pretty appropriate.

Some of the justification for antibiotic selection listed in our packet I don't feel was very rigorous. The Stanford

guide and the Johns Hopkins antibiotic guide were listed as justification for selection of a couple different agents. However, I know there's a Technical Advisory Panel that reviews selections quarterly, I believe.

So the scientific rigor I think is good. However, I do think there is a measure to harmonize and continue to harmonize with national guidelines as they come out, which I think CMS is doing. The work group thought it was very useable and feasible, as well, with compliance greater than 95 percent.

So, points of discussion, I assume we'll go into the bundle discussion, as well as whether this measure is topped out, and I'll turn it over to the group for discussion.

CO-CHAIR MORRIS: Anybody want to start?

MEMBER ZAMBRICKI: I would just comment that looking at the literature it seems like there's the closest correlation between this measure, selection of antibiotic,

and surgical site infection as compared to the timing measure, so I think this is a stronger measure than the timing.

MR. BRATZLER: This is Dale Bratzler. Can I [inaudible]10:37:01 the developer make a quick comment?

CO-CHAIR MORRIS: Sure.

MR. BRATZLER: So I'm just going to make a couple of quick comments about the three measures, and then I'll go back on mute, the three measures that are being considered, selection, timing, and discontinuation.

First, the measures all three are actually are strongly evidenced, and new guidelines will be published this year through four different specialty societies. All three measures are still strongly enforced in the new guidelines that are going to come out.

There still is opportunity for improvement in all the measures. I mean, if you look at national rates, they've gone up dramatically, but by different surgery type

there are some variations.

Finally, I actually would argue the point that was just made about the strength of evidence around antibiotic selection versus timing. I can't go into all the details right now, but there have been a spate of articles looking at different measures, trying to show whether or not they're associated with patient outcomes.

And I will simply say that most of those have some fairly large methodologic flaws, the biggest of which is they try to use performance rate published on --

(Inaudible due to telephonic interference)

-- to predict outcomes at the hospital level, something you really can't do, because that approach doesn't take into account all of the exclusions from these performance measures.

And then we're in the editing process right now of a very large study of all the SCIP --

(Inaudible due to telephonic interference)

-- for three years that shows you definitely want to go to a --

(Inaudible due to telephonic interference)

-- measures if you're eligible --

(Inaudible due to telephonic interference).

So I'll just say that there is strong --

CO-CHAIR MORRIS: Dale, your voice is kind of going in and out. Is it possible to make him a little bit louder? I'm not sure exactly why that's happening, but I just want to make you aware of it.

MR. BRATZLER: Okay. I'm getting big feedback sometimes when I speak. So I'll just -- I'll just end by saying there is strong evidence base for all three measures. All three measures will be strongly supported in new multi-specialty society guidelines that will be coming out this year.

Those societies include Infectious Disease Society, SHEA, Surgical Infection Society, and the American Society of Health

System Pharmacists. The four societies have been working now for the last couple of years on new guidelines. These three measures are all consistent and strongly supported in the new guidelines.

Finally, we have really good patient level data adjusted for the hospital effect, all the appropriate risk assessment methodologies that show that at the patient level you want to be in a hospital that passes these measures.

The measures are associated with improved patient outcomes, in contrast to some of the papers that have come out that have had strong, substantial methodologic flaws. So more will be coming on that soon.

CO-CHAIR MORRIS: Thank you. Paula, did you have something you wanted to add?

MEMBER GRALING: Well, my question was, again, where laparoscopy is in the exclusion criteria, and I think we've talked

around the table about with the trends in surgery that that's a concern.

MR. BRATZLER: I think, Wanda, you're on the call. Haven't we removed the laparoscopy exclusion?

MS. JOHNSON: For January 2012, it will be coming out.

MR. BRATZLER: So we've addressed it.

CO-CHAIR MORRIS: Could you repeat that?

MR. BRATZLER: We have addressed that issue. In January of 2012, the specification manual removes that exclusion.

CO-CHAIR MORRIS: Okay. Thank you. Any other comments, questions? Dr. Morton?

MEMBER MORTON: I think I speak in support of the measure. I think appropriateness of antibiotics is pretty critical.

I think the question about topping

out is probably still there and whether or not it fits within a composite, but certainly we want to give the appropriate kind of antibiotic, and before some of these measures were in place there was a wild, wild West about which kind of antibiotics were being used, so I think it's a very useful measure.

MEMBER COLLINS: You know, and a point of clarification, too, I was referring to individual agents within the submission here. I'm very happy to hear that the national, you know, kind of four group guidelines will be factored into the SIP initiatives for antibiotic selection, timing, and such.

I know those are coming out. I believe anticipated publication is around September of those guidelines, so, as a whole, I was a referring to just the individual choice of, you know, maybe a specific agent, not the overall choice.

CO-CHAIR MORRIS: Are we -- are we

ready for a vote? Okay. So, importance to measure and report. Eighteen say yes. None say no.

Scientific acceptability of measure properties. Fifteen say completely. Three say partially.

Usability. Sixteen say completely. Two say partially.

Feasibility. Fifteen say completely. Three say partially.

Does the measure meet all of the NQF criteria for endorsement? Eighteen say yes. None say no.

Apparently, we also should permit the STS and the ASC to introduce their measures. This is going to be -- this is a daily opportunity, so we'd like to invite the STS and then the ASC to introduce their measures before we proceed to them.

MR. JACOBS: I'm Jeff Jacobs from the Society of Thoracic Surgeons, and I think it's clear in the packet that there's three

antibiotic-related measurements that STS has put forward that are bold measures that have previously been reviewed. One is related to appropriate choice of antibiotic, one is related to timing, and one is related to the length of antibiotic usage.

There's extensive justification in the peer reviewed literature for all three of these measures, including manuscripts written by the STS Evidence-Based Task Force, that discusses the level of evidence to support each one of these three measures. I don't think I really need to go through those in great detail, because that's all supplied in the packet.

MS. SLOSBURG: Just to reiterate, yesterday we did try to harmonize with the SCIP measure for the IV antibiotic timing.

CO-CHAIR MORRIS: Okay. So the next measure is Measure 128, duration of prophylaxis for cardiac surgery patients. Dr. Kleinpell will be presenting this.

MEMBER KLEINPELL: Great. Thank you. This is Measure 128. It's duration of antibiotic prophylaxis for cardiac surgery patients. It's a maintenance measure. It was first released in 2004. The measure steward is STS.

With respect to this measure, we actually looked at the categories. Obviously, it is important with respect to prolonged antibiotics and the percent of antimicrobial resistance. With respect to scientific acceptability, I think our group had the most discussion about this because of the time line of 48 hours versus 24 hours.

Connie, unfortunately, she's not here, but with her involvement with respect to the Infectious Disease Society she actually brought forth some references for us. It was interesting to look at some of those with respect to the development of these guidelines for cardiac surgery.

It was developed originally by the

American Society of Health Systems Pharmacists, and they identified that expert opinion was a driving force with some respects with this time line of 48 hours. So we really questioned, going back and forth, 24 versus 48 hours with respect to the evidence for scientific acceptability of this measure.

Some of the other things we noted within it itself, there are some denominator exclusions listed, but we noted that they are -- their exclusions could not be captured in the previous version of the STS database.

However, it was indicated that the new cardiac surgery database, which was released this January of 2011, will enable exclusion data to be captured, and so that is forthcoming.

We noted that disparities of care were provided. For the most part, there still is a gap ranging from about 83 percent to 100 percent, but the mean is around 94 percent itself.

There are no direct costs. It's indicated with respect to the measure but really no specific information about costs involved in maintenance of this measure.

So I think for us the areas of controversy, again, was really along the timing and the evidence that supports that. Now, you did tell us that there was some forthcoming guidelines from the other societies, and I'm not sure to the degree to which they will give us more information on the issue of timing with respect to cardiac surgery antibiotic prophylaxis.

MEMBER COLLINS: You know, I can comment on that. The upcoming guidelines, which I have seen a draft recommend 24 to 48 hours, they say the evidence is inconclusive for one or the other, so they do allow that range.

MEMBER KLEINPELL: So, I guess, with that in mind, does this mean when we see it again next year that this might be changed,

then, to 24 to 48 hours, or can the measure developer speak to that?

MR. JACOBS: I think that there's very limited experience of doing cardiac surgery with only 24 hours of antibiotic prophylaxis, and the risks associated with an infection after heart surgery is probably worse than after most operations, because it's mediastenitis, which generally leads to death.

I think that prior to making a society-based recommendation of changing the length of antibiotic prophylaxis from 48 hours to 24 hours after cardiac surgery, I think that topic would have to be studied in greater detail in cardiac surgical patients. The evidence base in the literature simply does not support changing prophylaxis for cardiac surgical patients to 24 hours.

MEMBER HALPERN: And I have to say the same thing kind of exists for some vascular patients, you know, again because of the risks associated with being wrong is a

graft infection, and a graft infection can lead to limb or life loss, so we have a similar issue going on in vascular surgery patients.

MR. JACOBS: So, I would say it may be very reasonable to ultimately change this to 24 hours, but I think it's very premature to do that now, because it's simply not been studied.

MEMBER HALPERN: Also, I mean, you know, the reason for 24-hour coverage is that's theoretically when the skin seals, but if you look at like older people with very loose, kind of yucky skin, they don't seal at 24 hours, so it may be actually more patient-specific than has ever been looked at.

MEMBER CARPENTER: We'll get into this in another topic, but orthopedics is a similar problem with major implants and risk for infection and the 24-hour range. There is no data one way or the other, so it's been selected at 24 hours for majority of

surgeries, cardiac surgeries excluded from that for some reason, to a 48-hour window. You know, there's a question should we be consistent across the measures.

CO-CHAIR MORRIS: Any other questions or comments? Let's go ahead and vote on this measure.

MR. BRATZLER: This is Dale. I just one to make one --

(Inaudible due to telephonic interference)

-- guideline is going to explicitly recommend less than 24 hours for all operations based on no good evidence of prolongation being useful. Just a point of --

CO-CHAIR TORCHIANA: I'm not sure we heard that comment, Dale. Could you repeat it?

MR. BRATZLER: I said the new Multi-Specialty Society guideline is explicit of less than 24 hours for all operations, no exceptions.

MEMBER COLLINS: That must have

changed since public draft comments, then. Is that correct?

MR. BRATZLER: I don't remember exactly. It's been quite a while back that the draft was out there, but the guideline is explicit now that less than 24 hours for all operations. Actually, for almost all types of operations there are studies that have looked at single-dose prophylaxis.

You know, I think STS a number of years ago, they chose 48 hours, made that recommendation based on the fact that there weren't studies that compared 24 to 48, but they're actually, you know, limited, I recognize, limited studies of single-dose prophylaxis even in cardiac surgery.

I think our emphasis is much more in the guidelines on doing the up-front things right, correct dosing, correct antibiotic, redosing in the OR, which have all been shown in fairly good trials to reduce infection rates, and part of the big push on antibiotic

stewardship is to reduce unnecessary use of antibiotics.

CO-CHAIR MORRIS: Okay. We'll ask for the remaining comments about measures that are not their own to be held until the public and member comment period unless it's a member of the Committee and people who are the developers of the comments that are being discussed. Let's --

MR. JACOBS: Can I respond to that from STS?

CO-CHAIR MORRIS: Sure.

MR. JACOBS: I would just say that I want to reiterate what I said before that ultimately it may be appropriate to change cardiac surgical prophylaxis to 24 hours but that the evidence base simply does not exist in the literature.

CO-CHAIR MORRIS: Yes, you told us that. Thank you. Let's go ahead and vote on this measure unless anybody on the Committee has anything else that they'd like to say

about it. Okay.

Does the measure meet NQF criteria for importance to measure and report? Eighteen say yes. One says no.

Scientific acceptability of measure properties. Ten say completely. Six say partially. Two say minimally. One says not at all.

Usability. Thirteen say completely. Six say partially.

Feasibility. Eleven say completely. Eight say partially.

Does the measure meet all the NQF criteria for endorsement? Seventeen say yes. Two say no.

The next measure we're going to lump back up again to Dr. Collins, 126, selection of antibiotic prophylaxis for cardiac surgery patients, and this is the STS Measure 126.

MEMBER COLLINS: So, yes, this is a measure very similar to 0528 from CMS. This

is selection of antibiotic prophylaxis for cardiac surgery patients submitted by the STS.

The numerator first is the appropriate choice of antibiotic, denominator, number of surgeries. To meet criteria, patients must receive either first- or second-generation cephalosporin or vanco or a fluoroquinolone if there are allergies or contraindications there.

Much like 0528, the work group felt this was an important measure, scientifically acceptable. It's been active since 2007, so I feel it's both useable and feasible, as well. The issue of whether it's topped out could come up again at 92 percent compliance on this.

CO-CHAIR MORRIS: Anything else from the Committee, questions or comments for the developer? You made a comment previously that mediastenitis leads to death, so that's a pretty extreme sequelae of inadequate coverage, and so 92 --

So something, I guess, that we should think about in terms of being topped out, is 92 percent actually topped out for this measure? I guess in my opinion it's not, but I'd be certainly happy to hear anybody else's opinions about that.

Okay. Should we go ahead and vote, then? This is a quick discussion.

Does the measure meet NQF criteria for importance to measure and report? Nineteen say yes.

Scientific acceptability of measure properties. Fifteen say completely. Four say partially.

Usability. Seventeen say completely. Two say partially.

Feasibility. Eighteen say completely, and one says partially.

Lastly, does the measure meet all the NQF criteria for endorsement? Nineteen say yes. None say no. None abstain.

The next measure is Measure 0125,

timing of antibiotic prophylaxis for cardiac surgery patients, and this is going to be introduced by Ms. Zambricki.

MEMBER ZAMBRICKI: Hello, everyone. My measure is the percent of patients 18 and older undergoing cardiac surgery receiving prophylactic antibiotics within one hour of surgical incision or start of procedure.

The importance to measure, I think the summary of the evidence regarding deep sternal wound infection is very strong. Our group had quite a bit of discussion about is there evidence of a link between the measure focus and the desired outcome.

Is that link strong, because that's one of the criteria for importance to measure, and I think there is controversy about that. It sounds like there's a new study coming out, but the idea that exactly one hour before incision and one hour and ten minutes is not acceptable, one hour is

acceptable, I think there is some controversy about that in the field.

What is the exact timing that is necessary to have a link between decreased surgical site infection or not? And there have been some studies recently that show that that link is not a powerful link, but it sounds like there are studies coming out, so it's just a tough, tough issue I think right now, that importance to measure, for us.

The importance to measure section by the sponsor, for the rating of strength of evidence they left that blank and put "Not applicable." For the, "Is there controversy?" they left that not applicable, so they really did not address the issue of the rating of the evidence.

In terms of scientific acceptability, the exclusion laparoscopic is still there, but we understand that that is going to be taken away. It's interesting, because they do include patients for whom no

incision is required but make an exclusion for patients with laparoscopy, so that's -- it would be interesting to hear the explanation of that thinking.

In terms of the testing -- oh, I guess I should go back, the importance to measure. I think there's a real question about whether a gap exists for this measure. The median performance is 99.2 percent, and the mean is 98 percent, so it seems as though this should have consideration as far as being a topped out measure.

Testing, there was no data on exclusions, because there was some type of a database change with STS. They said that that would be provided in future years, and in terms of usability, there's a lot of similar measures, so the idea of harmonization is an important one.

I think that question of is there distinctive improved or added value, the supporters said not applicable, and I think

that gets back to the question about the controversy in the literature.

CO-CHAIR MORRIS: Anything else from the Committee in terms of questions or comments for our developer?

MEMBER ROGERS: I'm sorry, Christine, I didn't quite understand the issue of an hour versus an hour ten. Is the controversy with respect to that specific timing or some variation in timing?

MEMBER ZAMBRICKI: Well, I think the overall controversy is is there a connection between the SCIP measures and decreased surgical site infection for individual SCIP measures, and there is recently literature in JAMA and others with large patient populations -- at the VA there was 60,000, and JAMA I think it was 400,000 patients, or it may have been vice versa -- showing no correlation between individual SCIP measures, of which this is one, and decreased surgical site infection.

I would say in the field, and I'm interested in what the surgeons think, but in the field there is a lot of, I think, credibility problem with surgeons who get dinged because the case is delayed. Now they haven't met this SCIP measure and ask, "Well, does this really make a difference in infection if it's 15 minutes late or if it's one hour?"

So I think that's just kind of background noise in the field as far as the credibility issue, but in terms of the literature and the exact time interval, it's kind of -- there must be some time interval that's important, but it is not conclusive.

I know Connie's not here from Infectious Disease. She had presented us with some data that was quite dated, but it did suggest that 30 minutes might be appropriate.

So is it 30 minutes? Is it one hour? Is it an hour and a half? Is it two hours? It seems like there is controversy

about that.

MEMBER ROGERS: So there's no controversy --

MEMBER HALPERN: I would also say maybe even longer would be needed. We have a PharmD with us, but, you know, getting tissue levels, you don't have tissue levels for some antibiotics in an hour.

MEMBER ZAMBRICKI: It's not even an hour, because it's within an hour, so people sometimes will bring the patient in the OR, push it in, and it's gone in four minutes before.

MEMBER HALPERN: Right, and you haven't gotten tissue, skin tissue levels --

MEMBER ZAMBRICKI: Yes.

MEMBER HALPERN: -- because the volume of distribution may be -- and especially in the obese patients where the volume of distribution is high.

MEMBER COLLINS: Yes, I would concur. It's absolutely variable by

antibiotic, you know, and by procedure. The data I think is very good for extended periods of time. Administered very much too early or even within 15 minutes, there's harmful data there. It's difficult to -- you know, I see the discussion of exactly one hour versus one hour, five minutes.

I see why that is a discussion point. I don't know. I'm not sure of a better way to go about it, though. I think one hour has been studied, and I don't know if I have answers to this.

MEMBER MORTON: I agree with Curtis. I mean, it's a tough one to just figure out if one hour is exactly right or if it's 30 minutes, but you've got to set a goal at some point, and it's the goal that's been set, I think, through compromise. There's a lot of old data to show that, you know, the antibiotic does get into the tissue around that time.

I know that people are trying --

have to game the system to some degree, you know, as long as you do it before the incision, but it's a start. I don't know how else to do it other than to refer back to the developer and say that, you know, you need to have a better idea of when the antibiotic is actually getting into the tissue.

Regardless of those studies that came out, I'd like to hear from the developer, because I heard some mention earlier that there was some methodological concerns about it, that there were a lot of exclusions that were associated with those studies, and to get a better handle on what they think the scientific validity of those studies are.

CO-CHAIR MORRIS: Jeff, would you like to respond to that?

MR. JACOBS: So, what I'd like to do is just read two sentences from the measure submission form to clarify exactly what the recommended timing is based on specific antibiotics, and then I'll move from there to

addressing the question.

We say that "in patients for whom cefazolin is the appropriate prophylactic antibiotic for cardiac surgery, administration within 60 minutes of skin incision is indicated (Class I, level of evidence A)."

Then we go on to say, "In patients for whom vancomycin is an appropriate prophylactic antibiotic for cardiac surgery, a dose of 1 to 1.5 grams or a weight-adjusted dose of 15 milligrams per kilogram IV slowly over one hour with completion within one hour of skin incision is recommended (Class I, level of evidence A)."

And the reference for these two is a manuscript that was published in the "Annals of Thoracic Surgery" by the STS Evidence-Based Task force. In this particular manuscript, the first author is Rich Engelman, and this is a task force that spent a substantial period of time reviewing all the literature about antibiotic prophylaxis and came up with

consensus-based recommendations, including class and level of evidence.

So I think that these recommendations, they are, first of all, made on an antibiotic-specific strategy. Second, there is a level of evidence provided here.

Clearly, it wasn't filled out in the appropriate place on the form, but it was placed in this location, and this level of evidence came from an expert panel reviewing multiple manuscripts and multiple studies about this topic.

CO-CHAIR MORRIS: Can you tell us what was the year of publication for that citation?

MR. JACOBS: This is April 2007, "Annals of Thoracic Surgery", which is the most recent multi-specialty evidence-based medicine review of the literature that's been undertaken and published.

MEMBER MORTON: This supporting data is from an expert panel that reviewed the

existing literature?

MR. JACOBS: Everything that had been published up through that time, and the reference list to that is substantial. It's massive.

CO-CHAIR MORRIS: Any other questions or comments? Let's go ahead and vote.

MEMBER ZAMBRICKI: I would just ask one question, and that is the performance, 99.2 percent, and 98 percent mean, 99.2 percent median. Any thoughts on that?

MR. JACOBS: Yes, I think that if I was having heart surgery at a hospital that did 500 cases and I was one of the two patients that missed that and ended up dying of mediastenitis, I'd feel real bad. It's a very high-stake game, and I think that this 99 percent isn't acceptable for this, because if one percent of your patients die from this, it's a big problem.

CO-CHAIR MORRIS: Any other

comments? All right. Let's go ahead and vote. Importance to measure and report. Seventeen said yes. Two said no.

Scientific acceptability of measure properties. Eleven said completely. Eight said partially.

Usability. Thirteen said completely. Six said partially.

Feasibility. Fifteen said completely. Four said partially.

Does the measure meet all the NQF criteria for endorsement? Seventeen said yes. Two said no.

So now we'll move on to Measure 265, prophylactic intravenous antibiotic timing, and this was to be presented by Steve Findlay, but instead he has comments that will be read by Alexis? By Alexis. We'll also ask for the people who were present on that work group telephone call to please chime in even more than usual.

MS. FORMAN: Performance on this

measure has been above 95 percent for five or more years. On the Work Group D call there were questions about the requirement that antibiotics be given within one hour, rather than one to two hours or so.

Doubts about the strength of evidence for one hour, so discussion of that and questions to the steward should be prompt. That was the main issue.

Disparities data is not collected or available on this measure, so even though it looks topped out, since there is now a routine practice, there still may be a need to preserve the measure. Scoring on the measure for usability, science feasibility was all CRP in importance to measure and meeting the criteria.

The stewards, as they did yesterday, are likely to argue that this is a key measure for them. Personally, I think the measure is close to topped out, if not there already.

CO-CHAIR MORRIS: Any other comments from folks who were present on the work group call?

MEMBER KLEINPELL: Well, I know we did have some discussion with respect to the fact that there is no data on disparities, so I think that's something that we would look to have the measure developer provide if any of that data is possible.

We also had some discussion with respect to the definition and the issue that it says on time, and then the clarification of on time is within one hour of incision time, so it was questioned, well, couldn't that just be put into the statement? Instead of saying on time, say within one hour of incision time, because then it's clear, but that's semantics, I guess, but that did come up in our discussion, as well.

CO-CHAIR MORRIS: Okay. Would the ASC like to respond to that?

MS. SLOSBURG: We can definitely

add the one hour to the -- instead of on time if you think that's clearer.

MEMBER KLEINPELL: It makes it just more clear. Then you don't have to go to the subscript information and all that. So, great. Thank you.

MEMBER CARPENTER: In what ways is this different than the SCIP measure, other than it's outpatient?

MS. SLOSBURG: I don't know that it's different other than we look at, instead of procedure codes, we look at all patients who received an IV antibiotic prophylaxis. It's less burdensome for collecting data from ASCs.

And, again, to the issue of topped out, we've got about 900 out of the 5,200 reporting right now, and, again, right now we do not have disparity data because it is not mandated, but once it is, we will.

CO-CHAIR MORRIS: Anybody else want to comment about that?

MEMBER KLEINPELL: I think I would advocate, just based on that alone, that this should be retained, then, even though it is topped -- you know, the scoring is high. We just don't have that information on disparities, and you indicate that it's really a sub-portion of all the centers that are providing this data.

CO-CHAIR MORRIS: Okay. So we've asked for the developer to change the language a little bit regarding clarifying what on time means. Let's go ahead and vote on the measure.

Does the measure meet NQF criteria for importance to measure and report? Seventeen say yes. Two say no.

Scientific acceptability of measure properties. Ten say completely. Two say -- nine say partially.

Usability. Twelve say completely. Seven say partially.

Feasibility. Thirteen say

completely. Six say partially.

Does the measure meet all the NQF criteria for endorsement? Eighteen say yes. One says no.

The next measure is 0527, prophylactic antibiotic received within one hour prior to surgical incision. This is a SCIP measure being presented by Ms. Zambricki again.

MEMBER ZAMBRICKI: This is another in a series of antibiotic timing measures and also makes the exception for two hours for vancomycin. I'd say the only thing different with this measure is that the compliance is 97.1 percent, which is a little different, a little lower than the cardiac surgery measure.

So I think it raises the question of whether this is topped out or not, 97.1 percent, a lot of improvement in the last nine years. When it was first measured, it was 55.7 percent, so that's a lot of -- a lot of change. I think the questions regarding the

timing are still the same questions as they are for the other measures.

MEMBER HALPERN: Are the exclusions the same as in the other CMS measure?

MEMBER ZAMBRICKI: You know, I have to look back and see that.

MEMBER HALPERN: It said see --

MEMBER ZAMBRICKI: I think it still is the laparoscopic. I'll look at it while everybody's talking about other things.

MEMBER HALPERN: Well, the reason I ask, actually, because I missed this the first time I was reading the other measure on the prophylactic selection, is that they actually have as an exclusion those that did not receive antibiotics, so I wondered about that.

MEMBER ZAMBRICKI: Yes, these exclusions are length of stay greater than 120 days, hysterectomy and C-section, preoperative infections disease, performed entirely by

laparoscopy, enrolled in clinical trials, physician, advanced practice nurse, physician assistant documented infection prior to surgical procedure, procedures requiring general or spinal anesthesia that occurred within three days prior to or after the procedure interest, receiving antibiotics more than 24 hours prior to surgery, receiving antibiotics within 24 hours prior to arrival.

MR. BRATZLER: This is Dale. I can answer that question about that one specification. If the patient gets no antibiotic, they fail the measure on antibiotic timing, but because we assess, you know, antibiotic choice, they are excluded from the antibiotic choicer measure. So they fail one, but they don't fail both. It's to avoid double jeopardy on a measure.

CO-CHAIR MORRIS: Is that clear to everybody? Does that answer your question? Okay. Any other issues that anybody wants to raise?

MEMBER ZAMBRICKI: I was just wondering about the laparoscopic procedures for this sponsor.

CO-CHAIR MORRIS: Dale, has that been removed, and can you also address that question about whether this should be considered topped out?

MR. BRATZLER: Yes, so the laparoscopy exclusion has been removed, also, so for all of the SCIP measures. You know, I guess it depends on how you define topped out.

There is variation between surgeries, so, you know, you heard earlier that cardiac surgery has high rates of performance, but I can tell you the rates of performance for general surgery are lower. So it does depend on the type of surgery.

There is variation between types of surgery. I don't have the disparity data, though I actually think Wanda has provided it or can provide it, but there is some variation between different types of surgeries.

MEMBER COLLINS: I guess I would argue against this being tapped out for the reasons we've heard and the importance of this with the other procedures, as well.

CO-CHAIR MORRIS: Any other thoughts about that among the group?

MEMBER STAFFORD: I'll just say looking at the disparity data there is a fair amount of disparity based on age, geographic location, and things, so while the overall numbers might look topped out, it looks like there's a fair room for improvement in quite a few places.

In fact, you probably don't want to live in the U.S. territories looking at all the disparity data for the SCIP measures, so if you're going to Guam, be careful.

CO-CHAIR MORRIS: Okay, let's go ahead and vote on this measure. First of all, importance to measure and report. Nineteen say yes. None say no.

Scientific acceptability of

measure properties. Thirteen say completely. Six say partially.

Usability. Fourteen say completely. Five say partially.

Feasibility. Eighteen say completely. One says partially.

Lastly, does the measure meet all the NQF criteria for endorsement? We need one more vote here. Seventeen say yes. One says no.

Do you -- okay. So, there we are. The next measure is Measure 0529, prophylactic antibiotics discontinued within 24 hours after surgery end time, and this will be presented by Dr. Kleinpell.

MEMBER KLEINPELL: Right, 0529, prophylactic antibiotics discontinued within 24 hours after anesthesia end time with the exception that it's 48 hours for cardiac surgery.

It's a maintenance measure. It's been in use since 2001. The measured steward

is CMS. With respect to this measure, obviously we've indicated the importance in terms of antibiotic therapy before and discontinuing within an appropriate time line.

With respect to scientific acceptability, you know, again the issue is that there is evidence, obviously, about prophylactic antibiotics. Connie had raised the issue, however, that there is some but not a lot that single-does prophylaxis versus 24 prophylaxis, so again that whole time line is still -- we don't have the evidence to substantiate what is actually better for patients.

She actually indicated there should be a movement for no post-operative antibiotic prophylaxis, as there isn't evidence that supports any post-op prophylaxis, and again that's from her perspective from the Infectious Disease Society, but she's not with us to further expand on that.

So we did have some discussion with respect to that. We noted that the exclusions still listed laparoscopic procedures, so we wanted clarification on whether that was removed.

With respect to other aspects, we noted that the national average is 95.5 percent. However, there still is a gap, particularly with respect to disparities of care. It's about 88.7 percent with Hispanics, so therefore, obviously, it's continued importance in terms of monitoring.

In terms of usability, it currently is in use for the Hospital Inpatient Quality Reporting Program under CMS, and it's also part of the SCIP measure set.

We did have one issue with respect to feasibility. It's indicated that the specifications, which includes coding and data elements, are modified every six months, according to feedback provided by clinicians and hospital staff collecting the data, and so

we were wondering how this really -- how these potential modifications of specifications every six months is communicated to NQF and stakeholders and how it's expected that that may affect performance rates from quarter to quarter, so that was really our only other additional point.

Connie also highlighted there were no studies performed on cost of implementation. It was an additional comment that she had, as well.

CO-CHAIR MORRIS: Any other comments? Dr. Collins?

MEMBER COLLINS: I don't have many comments. I think we did a nice job of summarizing there. I think I would -- the 24-hour mark is pretty standard for this measure, and I would concur about looking at further data, shortening that duration when possible if the data is there.

MEMBER CARPENTER: Just from a practical standpoint, I think this is not met.

Most of the times this is not met it's because there is a small delay in administration of the antibiotic, so it's 25 hours. A lot of these drugs are Q-12, Q-8, so that last dose is right at -- if people give 24 hours of antibiotics, which frequently they do.

Maybe it's not necessary, but that's still built into a lot of post-op protocols. That last dose is right on the edge, and so a lot of the times this is not met because of a nursing administration, pharmacy delivery, patient availability issue, rather than a quality of care issue regarding this.

Now, maybe that should lead people to give, you know, 16 hours or 12 hours or something. That's another issue, but, you know, I don't think we should change it.

It's pretty standard now, but you wonder if 26 hours or something like that would have been a better, you know, time frame in terms of the numerator rather than strict

24 hours. Just a comment.

CO-CHAIR MORRIS: I think what you're describing here is more of an annoyance than anything else. Like Dr. Morton pointed out earlier, we have to have some target, and you're sort of saying that, too.

Any other comments or questions? Okay. So one of the -- one of the interesting issues for me that came up in the preceding measure was that we've seen a real shift in hitting the measure, from 55.7 percent to 97.1 percent over a long period of time.

One of the things that hasn't come up with some of these maintenance measures is have they had an impact? Have they -- it has come up for some, but not all of the measures have they had an impact, and it's certainly something that I think we should be paying attention to. Is this doing anything?

So there was a question about evidence of a quarterly change in performance, I think. Did that come from you? Okay. So

we'd like to ask CMS to address that and also the laparoscopy question, as with the previous measures.

MEMBER HALPERN: And I think it also goes back to what Christine was saying. Are we actually affecting the end point, which is the patient?

MR. BRATZLER: So this is Dale. I'm not sure I completely understand the question, but across the board the laparoscopy has been issued January 2012, so it's out of these measures.

Remember that this measure is not about impacting infection rates, because when you look at all of the published studies that have compared short duration antibiotic prophylaxis for long duration or single dose to long duration, the outcomes for virtually every single one of those studies is the same, that surgical site infection rates are no different.

Prolonging antibiotics doesn't

lower infection rates. It certainly doesn't increase them, but it doesn't lower them. This measure is about antibiotic stewardship, stopping the use of unnecessary antibiotics.

So, in terms of measuring an outcome, it's one of the flaws in some of the studies that have been published that I mentioned earlier that are looking at composite SCIP measures. This measure has never been shown in clinical studies to impact infection rates, and so it doesn't make sense to have it in a composite that's looking at, you know, reducing infection rate.

So this is, I think, maybe one of the best performance measures that we've ever had in the country related to antibiotic stewardship, because, as noted earlier, the national performance on this has gone from 55 percent to 95 percent over about eight years, a dramatic reduction in unnecessary antibiotics.

MEMBER ZAMBRICKI: I have a

scientific question. I agree with you. This is about antibiotic stewardship, and I was wondering. Is there evidence that there is benefit to reducing surgical site infection with any antibiotics once the wound is closed?

MR. BRATZLER: Well, I would argue that -- so there is some experimental data that showed that there is a short period of time after surgical closure that the wound is vulnerable, so I never argue with the surgeon if they want to give, you know, a single dose at the end of the case. You know, I think there's a theoretical concern that wound contamination could occur.

Beyond the immediate post-operative period, I've never seen any studies that show convincingly that you can reduce surgical site infection rates by continuing antibiotics, and yet we know from a number of studies that you can increase the rates of C-difficile colonization.

Then the cardiac surgery

literature at 48 hours, the only paper that looked at 48, less than 48 versus greater than 48, when an infection occurs in a patient, it didn't -- I want to highlight again the infection rates were no different less than 48, greater than 48.

Infection rates, surgical site infection rates were no different, as has been shown in all studies, but when an infection did occur, those patients that had received more than two days of antibiotics had 60 percent increased risk of a resistant organism. You just select out, colonize the patient with resistant organisms.

So we strongly think that this is an important measure, and we're actually -- I think of all of the accomplishments of SCIP, this may be one of the most important.

MEMBER HALPERN: How does -- how does -- you know, in vascular surgery we have a lot of procedures that are not 100 percent clean, because there's like a gangrenous toe,

and I have to say I always get -- coding those patients appropriately in terms of clean contaminated, contaminated, I don't know where they fall in, and how does -- how do those kind of cases weight into your 24-hour antibiotic rules?

MR. BRATZLER: Yes, that's a real good question. So if there is documentation of infection before, during, or after the operation within 48 hours for most operations, 72 hours for cardiac surgery, the case is excluded from this performance measure.

Wanda can correct me if I'm wrong, but I believe gangrene that's documented actually is considered documentation of infection, recognizing that sometimes it's just ischemia, but those cases would be excluded.

MS. JOHNSON: I believe that's correct, and I know we've gone round and round with abstractors on this that gangrene is considered an infection.

MR. BRATZLER: So the case is not included in the denominator for this measure.

CO-CHAIR MORRIS: Okay. Any other --

MEMBER MORTON: I think he makes a great point. This isn't about preventing surgical site infection. It's about appropriate use, and I think it's pretty critical, because we've seen C-diff rates go up over the last ten years, so it's an important measure from that point of view.

CO-CHAIR MORRIS: Okay. Do you want to add another comment to that?

MEMBER COLLINS: From an antibiotic stewardship standpoint to at least, you know, in our institution this has had a big impact, and I feel it will and has nationwide, as well, really cutting back on durations of therapy, which, you know, can extend for too long, which puts patients at risk. So I concur with the impact that this has on antimicrobial stewardship and

stewardship practitioners across the country.

MEMBER STAFFORD: Yes, and I would say, not only that, it's about resource utilization and cost, so you've got nursing administration cost. You've got pharmacy preparation cost, extra tubing, all of that the healthcare system somehow some way pays for, and so, irregardless of the other issues, that's also one of the big issues with this, and that's another reason that this is a really important measure.

CO-CHAIR MORRIS: Okay. Let's go ahead and vote. Does the measure met NQF criteria for importance to measure and report? Nineteen say yes.

Scientific acceptability of measure properties. Fourteen say completely, four partially, one minimally.

Usability. Eighteen say completely, one partially.

Feasibility. Sixteen say completely. Three say partially.

Does the measure meet all the NQF criteria for endorsement? Nineteen say yes.

Next, we'd like to open the floor to NQF member and public comment, and I thought that we had given some time for member and public comment last night, but apparently there were some -- there was a member that would like to comment about the discussion from yesterday, as well.

Anybody on the phones for member and public comment?

OPERATOR: Star 1 to signal. We have Barbara Rudolph, I believe.

MS. RUDOLPH: Yes, hello? Can you hear me?

CO-CHAIR MORRIS: Yes, we can hear you.

MS. RUDOLPH: Okay. Thank you. I wanted to make a couple comments. The first one relates to some of the AHRQ measures that were voted down. I just want to remind the Committee that these measures are being widely

used. At least about 20 states, these measures are being reported publicly, and --

CO-CHAIR MORRIS: I'm sorry, could you speak a little bit louder?

MS. RUDOLPH: The measures -- I wanted to talk a little bit about the AHRQ measures that were turned down yesterday in the vote and just sort of remind folks that these measures are being widely used not only by state health data organizations that publicly report information but also by state Medicaid programs and state public health programs who rely on the state hospital discharge data and utilize the AHRQ measures.

So I just -- I think it's important to think about the widespread use of these measures and the fact that most entities that public report to day only have access to state hospital discharge measures data sets.

So when you think about things like 30-day mortality, you're really limiting the ability of entities to measure, because

the only party that can actually do that is Medicare, because they're the only ones that have enough cases to be able to actually use that 30-day measure and where they have information, because they have the longitudinal data.

So I would strongly urge you to consider not merging and not pairing the 30-day and the inpatient failure-to-rescue measures, because that would essentially mean that there would be no public reporting, so please consider the use part, as well as some of the other components.

The second -- my second comment relates more to the continued use by CMS of the clustered hierarchical models where they're using random effects as the estimation, as opposed to fixed effects, and there's been recent research that shows that when you place --

When you use those kinds of random effects models, you essentially focus on

specificity, rather than balancing sensitivity with specificity, and the difference there is you end up protecting hospitals and not providing any good information for either consumers or purchasers.

If you look at the CMS reports, you'll see that out of 4,500-some hospitals, and you have five or six or seven who are above average and an equal amount who are below average, and that really gives absolutely no information, and I think it misleads the public in terms of saying that these hospitals are okay, because they've been pulled to the average by the estimation techniques.

So just something to think about. The article, most recent articles by Kipnis, Escobar, and Draper, and it's in medical care, the May 2010 volume. So those are my comments for today.

CO-CHAIR MORRIS: Thank you for your comments. A couple of things that I

wanted to just clarify. One of the measures that you discussed, Measure 0351, which was the AHRQ death measure, we discussed that adjacent to a discussion of the failure-to-rescue measures.

Just so that the Committee understands, we did not vote that down. We voted in support of that. I'm sure that this will come up for further discussion. We voted in support of the AHRQ death measure.

MS. RUDOLPH: I was actually addressing the competing, the discussion for the competing measures at that point.

CO-CHAIR MORRIS: Okay.

MS. RUDOLPH: I know you had discussed yesterday potentially pairing them or just harmonizing them to the 30-day and eliminating the inpatient. So I would be very, I guess, cautious about doing that just because, as I said, there's so many groups that are currently publicly reporting those who don't have access to 30-day, and you're

also limiting the population then. Medicare only addresses the over-65, so for commercial payers and others the inpatient measure is very important.

MS. MURPHY: Barb, it's Melinda Murphy. The discussion about pairing or linking them was a discussion that was made by the developer. The Committee has not had that discussion, and it will have -- as you said, the three measures will be discussed whenever there is the discussion of related and competing measures.

MS. RUDOLPH: Okay. Actually, I have one more comment. When, you know, there are competing measures -- for example, I don't know if you're going to discuss the survival predictors or not, but I really wasn't aware that there was going to be this discussion, and we don't have our developer ready, so if there's a possibility, if that's going to be discussed, I'd like to ask for an extension.

MS. MURPHY: Related and competing

measures with respect to those measures that have been discussed yesterday and today will not occur today. Helen's going to give us some background and a bit of setup for that, and then we'll actually convene a conference call to talk about related and competing measures, and all of the affected developers will be notified in advance and invited to that meeting.

MS. RUDOLPH: Okay. Great. Thank you.

DR. BURSTIN: We'll start the discussion today, Barb, but certainly the detailed discussion we'll have on a subsequent call. We'll make sure you have sufficient notice to get your developers in line. The one measures we will talk about today, because the developers will be here this afternoon, are the pediatric heart surgery measures.

MS. RUDOLPH: Okay. Okay, great. Thank you.

CO-CHAIR MORRIS: Is there any

other public and member comment? Helen, would you like to frame out the discussion of the related and competing measures that will occur after lunch?

DR. BURSTIN: Sure. So, briefly, we do still have those two additional measures. Patrick will be back this -- we have not done the post-operative wound dehiscence yet, right? There's two additional measures still to do. Patrick will be back this afternoon. I emailed him to see if he can get back sooner.

But the other thing we'd like you to do today is actually just at least look at the measures that are before you, and I believe you have a table that outlines for you the measures that at least we would consider related or competing.

We also provided for you guidance that we've put out for comment recently specifically on our guidance of how you would even begin looking at relating and competing

measures and so would turn your attention to those documents. Is it easily findable on the thumb drive?

MS. MURPHY: It would be in the materials for the May meeting. It'll be in that PDF.

MS. FORMAN: It's the materials that went out on April 26 with the eight attachments, and it's Attachment 5, and I'll put it up here on the screen.

MS. MURPHY: So the Surgery Steering Committee materials PDF that has nine attachments, and Attachment 5 is related and competing measures. That's the table. It starts out with the memo. You scroll down. You can see the table.

DR. BURSTIN: Great. So this is basically just offering you an approach, and, again, this is not a new criteria, and over the last couple of years we have always had committees do related and competing measures.

The two big differences are, one,

for the first time you will actually, because of this new endorsement maintenance process, have all the measures put together in the same project, so you have the chance to look across the measures and see whether one measure is superior, whether another measure is superior.

But the key thing, also, is when there are more than one measure in a given area, when is it okay to have them coexist? When do they need to be harmonized, and if so, what's involved in harmonization? So we've tried to just give the Steering Committee just more guidance overall to specifically help you with that decision-making.

Could you scroll down to the part where you get to the nice flow chart, Alexis? Keep going. All right. So if you could just stop right there for a second.

So the first question is determining whether there's actually a need to see if there is assessment of competing or related measures. We've kind of done that for

you by laying out for you which measures we believe address, as we've defined it, competing measures, the same concepts for the measure focus, meaning the target population, process, condition, event. The example, you know, we've talked about several of these like failure-to-rescue over the last couple of days.

If you then go down to the next -- pull up that page there. Perfect. So here's where we really look to your input and your insight, so we would ask you to --

All these measures that you're going to be looking at have now been deemed as being ones that meet all the NQF endorsement criteria. So that's the first lens.

We don't get to related or competing until you get through that first step. The key thing then will be to look at some elements of each of the criteria to help you make that first assessment.

So the first one on impact,

opportunity, and evidence relates to the importance to measure and report, and in this instance we would see if, for example, there's any differences there. Is one measure superior, for example, because it may provide a different, a broader patient population, a bigger opportunity for improvement, for example?

Scientific reliability and validity, we could actually pull up your ratings to see if, in fact, you think there were differences and one being more highly reliable, valid, and precise than another one. And then, all else being equal, our preference would be that when we can, we want to get to the measures with the broadest possible population who could be measured.

So, there are some instances, for example, where there are measures that may be Medicare only or measures that may be, you know, a very narrow population. The preference would be a measure that allows the

maximum number of people to use the measure for reporting.

Usability, this is an important one, because this is where it actually gets at how usable it is for the accountability functions. Can people actually get those data and use them for public reporting?

That was the exact issue Barb just brought up for us on the telephone, for example, that, you know, in their experience the AHRQ measure is more usable, because the end audiences have had access to that measure, have used it for years for public reporting, and have found it useful, in addition to how useful is it for quality improvement.

That's where I think a lot of the discussions you've been having today about this is actionable, this isn't actionable would come up, and we'd ask you to take a look across the two of them.

In terms of feasibility, if there's two related measures and one has

significantly less burden on the data collector to do it, that would be one to consider, but, again, we want to try to move towards where we know the puck needs to go.

So we want to move towards measures that will get us closer towards EHRs and sometimes, for example, a claims-based measure may work now, but maybe a registry-based measure could work in an EHR environment to follow. So it's not a clear-cut split, meaning claims always wins here.

Then, finally, if a competing measure doesn't have clear superiority, and, unfortunately, I think a lot of the times we wind up being in this bottom bucket, this is where we would ask you to look to see if there's a justification for multiple measures in a given topical area and really consider whether the added value of having those two measures offsets the potential burden, the potential burden in terms of confusion, the one you guys talked about earlier in terms of,

"I do this in a hospital. I don't do this in the outpatient side," but also to get at the issue of whether you then wind up with confusion.

If people have different scores on different measures or if you think about this moving towards a more, you know, high-stakes payment model, would you have clinicians and hospitals ranked differently depending on the measure that's selected?

So those are the kind of pluses and minuses that we would ask you to weigh, and in general one of the things we've tried to do is point out that when we can -- you're fine, actually -- we would like to try to get to the measure that gets us to, as I mentioned, the broadest possible population, and when you want more than one measure, justifying it.

So, again, one thing to consider is you may have a registry-based measure in front of you. You may have a claims-based

measure in front of you. You would need to consider, first of all, are they equivalent? Are they really rising to the same level on each of those criteria?

And then, at the end of the day, do you see added value that you could justify by saying, "The world as it is right now could actually live in this world of having both of those and justifying it?" So that's what we'd like you to think through, and that's it.

So we'll walk you through this. Again, this is very new. We are -- we literally just closed comment on it about a week ago. You guys are forging a new path for us, as cardiovascular did, as well.

It's not a new criterion. It's really just that we're trying to give you as much as we can, sort of some decision trees to standardize our work across committees, sort of like the exercise we went through on competing, which we're now writing up the flow chart for based on this discussion this

morning.

So, that's what we'll do. We're not going to get through all the discussion today. I think what we'd like to do is for the competing measures, particularly for the developers who are here this afternoon around pediatric heart surgery, to hear from the developers the differences between the measures.

That's a little bit of a unique situation. We had two measures that came through, a pediatric heart surgery project that we just did, one from Children's Hospital Boston, one from STS. At the end of the day, the Steering Committee couldn't make an assessment of which they thought was superior.

It went to our CSAC, and their feeling was, "Well, the third measure that's competed is in your committee," so this would be the logical place to have this committee not have to reevaluate the first two measures -- that's been done -- but to at least give an

assessment of how the measures look when they compete head-to-head.

Is there a consideration for having more than one of those, more than two of those? Are there questions you would pose back to the developers?

We're not going to make that -- ask you to make that decision today, because we've just provided you those materials in the past week, but at least as we have this discussion to have you lay out what are the key questions you would want to ask the developers.

Is there additional information you would like the developers to come back with, and at times are there things that two of the developers could do together to bring their measure together, for example, as another option? So I'll stop and see if there's questions. Yes?

MEMBER CARPENTER: The term "harmonization" has been used multiple times,

I think for different purposes sometimes, and sometimes we think, well, that takes two similar measures and put them into one. I don't think that's how this is being used here.

It seems that it's more looking at the definitions, the numerator, denominator, the data that's collected and trying to make that as similar as possible, in part to relieve the data burden. Is that -- how should we be thinking about the term "harmonization"? How should we use that in our discussion?

DR. BURSTIN: It's an excellent question. When we talk about -- when we're talking about harmonization, we're talking about where there is the same measure focus for perhaps different patient populations.

So an excellent example is the exercise you just went through about the ambulatory surgery environment versus the inpatient surgery environment. They've got

different data sources.

It's got to be by definition a different measure to allow us to capture that data, but at the end of the day we've got to be making sure the science is consistent, the way the measure is constructed is consistent. That's harmonization.

At times, we do actually have examples of when two measures can, in fact, be put together into one. We just did this recently. It's somewhat painful, takes a long time, but, for example, we had a measure that already existed of doing cervical spine films for patients with trauma, what the indications were.

We then had a measure submitted that said CT scans for patients with cervical trauma. That's like, "Wait. This will create the wrong incentives out there if there's two," so instead those two developers decided to work together and actually came back to NQF about six months later with a combined measure

of CT plus cervical spine and MR.

We've been doing some work, for example, between the American College of Surgeons and CDC on two competing surgical site infection measures which, given how high-profile SSIs are, the idea of sending out to the universe two competing surgical site infection measures just did not seem optimal, so they have now been working for more than six months on trying to bring those measures together. So there are examples of both.

When we're talking about competing, we're really saying same patient population, same process of care identified. Sometimes they're on different data sources, different data platforms, which might be one reason to consider.

And even if you decide to put through competing measures and say there's justification for both, if they're really on the same population and they're really looking at the same measure focus, that's where

harmonization comes into play.

We don't want things defined slightly differently so that you wind up with apples and oranges, even though we know we may get very different rates of performance when you change the data platform, and I think we're just going to have to live with that for the next X number of years until we all move to the, we hope, the electronic, interoperable electronic platform.

CO-CHAIR MORRIS: I think it's time for lunch, then. Right now it's 11:55. Is Patrick going to be here at 1:00? Okay. We want him to be available. Let's -- no response yet, right, from him?

DR. BURSTIN: I can email him.

CO-CHAIR MORRIS: Okay. Let's -- okay, let's say 12:45 we'll come back. Everybody get that, 12:45? Okay.

(Whereupon, the foregoing matter went off the record at 11:56 a.m., and resumed at 12:45 p.m.)

CO-CHAIR MORRIS: Okay. It is time to start back up again. We have two more measures to discuss prior to moving into the next part of our meeting, and this is a continuation of general prophylaxis and wound dehiscence. The next measure is 0367, post-operative wound dehiscence, to be presented by Dr. Cima.

MEMBER CIMA: So, these are -- there are two of them. I'll just discuss them both together, because they basically come off -- it's basically the same template. It's just one's a pediatric -- associated with pediatric patients. One's -- the other one is adult.

Going reverse and back in the order, feasibility and usability, no one had any questions about that. I'm glad Dr. Romano provided an updated reference.

It is clear that using administrative databases that you're able to identify these cases quite readily and that

they truly are what they are, by and large, for the most part, what they say they are. They are wound dehiscence.

So usability and feasibility I think no one had any concerns about. Where we really got -- where the real issue hit the road was whether it's a measure to be reported and also the fundamental underpinnings of reporting it.

What I was most struck by, and I hope this is simply a clerical error and not an attempt to use the data in a way that may not support it, the summary of evidence of the reason for doing this is referenced throughout the paper on one paper from 1989, and I just want to read it.

It says, "Based on a two-stage review of randomly selected deaths, Hannon et al reported the cases with the secondary diagnosis of wound disruptions were three times more likely to have received care that departed from professionally recognized

standards, in cases without the codes, 4.3 percent versus 1.7 percent after adjusting for patient demographics, group geography, and hospital characteristics." Basically, that is his rationale for doing this.

Given the profound nature of that, I actually had that paper pulled, and, unfortunately, there is no separate analysis based on wound dehiscence. The actual data is based upon cases with a infection and/or wound disruption reported as a secondary diagnosis.

So there is no separate analysis based on wound dehiscence. Therefore, the statement there is incorrect. Furthermore, when they say that adjusted from recognized professional standards of care, a quote from the paper's authors, "In addition to the three targeting criterias that were not significant for other quality of care judgment, wound disruptions or infections was also not significant," so the paper's authors also said.

So what this really led us down, the feeling was the justification for this, the scientific justification and the whole rest of the thing that flow from it are not supported by it. Furthermore, further analysis and also in the paper provided by Dr. Romano talking about preventability of the event, in a review of the literature basically the vast majority of the literature would say that this is a non-preventable event.

It's related mainly to non-modifiable factors such as morbidity, AIDS, diabetes, obesity, underlying live function disease, and in this paper Dr. Romano so kindly provided this morning, if you look at the post-op wound dehiscence rate, non-preventable ranks as 25 percent of them. Uncertain of what caused it is 41 percent.

So that's consistent with what's in the literature of somewhere between only 20 to 25 percent of cases have anything that they can say is a possible modifiable factor, so

the scientific rationale for it and everything, especially the rationale that it's a huge standard of care problem is not supported by the literature.

As far as the demonstration of performance gap, looking at it over the extensive amount of data that was provided, the estimate is anywhere on range, particularly about one to three events per thousand with it being higher in the older age population, which is consistent with what we see in the literature, but it's certainly not a modifiable factor.

There is no real disparity data that is cited and provided by the developers, but there is no real difference in disparities. Again, it's mainly tied somewhat into age and possibly underlying diseases.

The data cited on disparities mainly goes more to the methodology, as opposed to specific to wound dehiscence, so the methodology just shows there is

differences in PSIs based on disparities, but there's no specific data related to wound dehiscence.

Then, finally, other than this updated version the data, there's some issues about what's in, excluded and in. This is supposed to be only abdominal pelvic surgery that is reported, and in the document as presented there's things like inguinal hernia repairs and stuff like that are presented, which really would not -- would dilute the things, but I think more recent analysis of the PSI and some of the improvements that have been made to it clearly show that it's a valid measure and a reliable measure.

So really the main issue just revolved around the scientific underpinnings here, the rationale for it. Is there opportunity for improvement? But the other ones, certainly the methodology and everything is top notch. So that was the -- and the only difference is one's a pediatric and one's not.

CO-CHAIR MORRIS: Anybody else on the Committee with questions, comments about this measure?

MEMBER DUTTON: Sure. I have one comment, one question. The fact that right now we think that most of the causes of this are unpresentable doesn't necessarily mean we shouldn't measure it, in my opinion.

Fifteen years ago, we thought central line infections were unpresentable and so on, and there's countless examples of that, so measuring that might provide an impetus to find ways to prevent it, but then the more specific technical thing, I spent many years working in a trauma center.

We do a lot of damage control procedures. We deliberately leave abdomens open on a fairly frequent basis and then obviously come back and re-close them at some later point, which as far as I can tell from reading the numerator and denominator statement here, those would count as misses on

this, and that's incorrect.

MEMBER CIMA: That, actually, is a coding standard, and it depends on how it's coded in the institution. So if the institution knows that it was a patient that was left open, they don't count -- they don't code it the same way.

I explored that with our coders, and they said there's a way of differentiating that, although that was one of the initial issues with the PSI-14, from the way I understood it, that coders were not distinguishing that, and people who were intentionally left open such as cardiac cases where the chest was left open, those were being coded.

That's been apparently remedied. I can't speak to that 100 percent, but my understanding is there is a way of differentiating that in the codes.

MEMBER DUTTON: Thank you.

CO-CHAIR MORRIS: Peter.

MEMBER DILLON: I think the points that you brought up in general about the validity of this as a quality measure are crucial to understand, so let me just make a few comments about 6-7, and realize that I'm, you know, I may well be biased in my interpretation of this, but we don't know that this is a problem in the surgical care of children at all.

There is absolutely no literature, and what I'm worried about is it's a classic case of having cited that article, as you said, which was just sort of cookie-cutter all the way through.

You know, it's a case of children are not small adults, and to take or to extrapolate an article from the adult literature and say it's a problem in the pediatric surgical literature is -- or in the pediatric surgical care I think is totally inappropriate.

In addition, this is an incredibly

small complication, and that's the problem. I don't think you're going to -- we can't even pick this up in our NSQIP data right now in order to be able to discern performance differences within institutions with the incredibly small numbers that this is going to be at least pertaining to this one metric.

So, as I said, I'm concerned about the small numbers, and it's also a tremendous -- it's going to be a tremendous data burden. The denominator is way too big. They've got way too many procedures included in it, so I have serious concerns about the validity of this metric right now as a quality measure in children's care.

CO-CHAIR MORRIS: When you say a very infrequent event, are you referring to pediatric, adult, or both?

MEMBER DILLON: Pediatric.

MEMBER ZAMBRICKI: I would like to comment on the frequency. Even in the adult literature that was accompanying this measure,

the rate was less than two per thousand, and out of that one of the studies identified 66 percent of those were not preventable because of patient conditions.

So I'm not sure you could say the 998 that didn't have it were providing quality care because they didn't have a dehiscence. It's just such a small number of occurrence, and then over half of those aren't preventable and aren't an indicator of quality, so one in a thousand.

CO-CHAIR MORRIS: I'm a little concerned about that number. I think that it should actually be a lot higher, which makes me wonder if the capture is inadequate here. Any of the other surgeons or clinicians want to speak to that?

MS. DAVIES: This is Cheryl Davies, one of the developers. If I could just address one point on the 66 percent for the preventability on the pediatric measure.

MS. MURPHY: I'm sorry, who is

speaking, please?

MS. DAVIES: I'm sorry. This is Cheryl Davis from the RQI team.

MS. MURPHY: Okay.

MS. DAVIES: So that study was done with a previous definition, and so wanted to just note that that also included some staged procedures, and that's why they were included in the non-preventable category, and since then we have now excluded some of those staged procedures.

Now, that being said, we don't have enough detail about that study to know exactly how that would change the preventability ratings, so we're not able to give an updated number for that, but we suspect that a good chunk of those were staged procedures that are now excluded from the indicator.

MR. ROMANO: Could I address the other questions? I'm sorry.

CO-CHAIR MORRIS: Sure, but

anybody else on the -- so I asked a question. Any other clinicians find that number to be a bit low, two per thousand?

MEMBER SIPERSTEIN: A comment. At our institution, I would agree with you that our reported number is less than what everybody clinically thinks the incidence is, so I think the true incidence is probably higher, and it's a vagary of coding, at least at our place.

MEMBER DUTTON: I deal with poor protoplasm. I'm certain our incidence is higher, but we also have a lot of procedures I'm not sure how you'd score. You know, the wound partially opened at the bedside, packed, that kind of thing.

MEMBER CIMA: There are some classifications on this. This is fascial dehiscence, so not wound disruptions and stuff, and that's where one of the initial coding issues was was wound disruption versus fascial dehiscence. This PSI is fascial

dehiscence that requires re-operative closure.

So I personally looked at every single one of these in my institution and then kind of correlated it with other measures, our OR schedule and stuff, and found it to be very close, and our rates are about this number.

If you make it -- if you use that definition, wounds that are opened in the ICU and left open are not included in this, this PSI, so it may be, you know, vagaries of the coding, as Allan said, but I don't see this number being much higher. At least, in my experience, it's higher than what this is.

MEMBER STAFFORD: Yes, I would agree with that. I think there is dehiscence, and there is dehiscence, and, you know, I mean, it's not uncommon for the fascia to separate a little bit. You have a little bit of fluid. You never have to do anything about that. It's the ones that you have to take back to the operating room that are probably the ones that are more actionable when you

think about it.

That being said, especially if you're talking about an elderly population, the risk factors for dehiscence are going to be a lot higher, as well, so, I mean, I think if you use a very strict definition it's probably more reasonable than --

I think, Arden, what you're talking about is the number probably is higher when we think about just some dehiscence, but even at the fascial level, but it's the actual number that you actually have to re-operate on is a different subject, and that probably is more the smaller number.

MEMBER MORTON: I agree with Barb. We look at all of these as part of our Professional Practice Evaluation Committee. We look at all the PSIs, and we actually look at these when they come up, and it's been pretty consistent. What the PSI reports is what we see in practice.

CO-CHAIR MORRIS: So I guess the

question, then, is is this measure -- are there enough events to actually measure with this? And it sounds as though the thinking is that they're not necessarily preventable. Those few events are not necessarily preventable.

DR. BURSTIN: Just one overarching comment here. These are intended to be patient safety indicators, so they are frequently very rare events, so there's not necessarily -- when it's a safety issue, I don't think there's necessarily a threshold for how many is enough.

Again, it's a claims-based measure, so it's not something for which there is a lot of data burden to collect, and the question is, for each of those events, you know, what are the downstream implications for that patient, and are there learnings that happen at that hospital as a result of knowing that, the way you --

MEMBER CIMA: The issue is a

safety event implies that you can alter the course by doing something. The vast majority of these you cannot, so is it a valid patient safety event? I mean, that's why one of the complaints against this one is that it is an event.

Is it -- should it say that an institution that has an event is not safe? That's not the point. We're not asked to do that here. We're asking is this a quality measure that's important to measure and report, and is it scientifically valid?

MEMBER ZAMBRICKI: Just another point. If I recall from the demographic data, there was a slight increase in the rate in academic medical centers, which I think brings up the point that the patients, many of us could predict what patients are going to dehisce, you know, people with bad tissue problems, vascular, just sick patients.

I think a lot of those patients get sent to academic centers, so it's possible

the better surgeons take care of the patients and can actually have a higher rate, because they are getting the referrals of these sick patients to their facilities.

MR. ROMANO: Could I address?

CO-CHAIR MORRIS: Sure.

MR. ROMANO: This is Patrick Romano representing AHRQ. So I think -- I think I can address almost all of these comments. So, first of all, the data burden is zero, because the data are already collected in the routine course of hospital activities.

In terms of the coding issue, yes, this code is specific for wound disruption, so if a hospital is using it in cases that were left open deliberately, that's incorrect coding. In the paper from the VA -- it's at your seats -- you can see that there was a positive predictive value of 87 percent, a confirmation rate.

When you look at the false

positives, seven out of those 15 false positives were cases where the abdomen was intentionally left open, so there is some degree of miscoding. What we found, of course, is that as coders learn how to use the codes and attend to it that this problem goes away.

In the pediatric study, I call your attention to Table 2, and it's a little bit tricky to read, but the numbers for post-operative wound dehiscence, there were a total of 102 cases in the children's hospitals that participated in this study, but only 52 of those were reviews.

So for three or five percent they found a coding error. That's in the third column of numbers, and then in five cases or ten percent they found that the event was actually present on admission. So, again, it's about 85 percent that were confirmed.

Now, then, to the preventability issue. So they reported that 34 percent of

the events were preventable according to retrospective review of the medical records in the pediatric study.

In the adult study from the VA, they don't specifically assign a percentage, but they do report on page four the reasons for the wound dehiscence, 32 percent fascial tearing, 12 percent necrotic fascia, clearly a patient factor, 11 percent breakage of suture material, nine percent intra-abdominal infection, two percent unraveling of sutures.

They further report that about 70 percent of those fascial closures were done by trainees. Unfortunately, we don't have a comparison group of patients who didn't have dehiscence, so we don't know what the comparable percentage of fascial closures that were intact done by trainees is. So there's limited data on preventability.

However, this estimate of 30 percent or so is in the same ballpark as for other risk-adjusted outcome measures, so if

you look at risk-adjusted mortality measures, the risk-adjusted complication measures that were discussed yesterday, I think that the empirical literature suggests that something between 20 and 50 percent preventability is the typical range that we see for risk-adjusted outcome measures. That pretty much comes with the territory.

In terms of the issue of importance, the importance of this indicator as it was originally endorsed by the NQF was based not on the prevalence but on the impact, so the average case, because of the tight way in which these cases are defined based on return to the operating room, the average case that experiences this complication in two fairly sophisticated case control analyses incurred an extra nine days in the hospital, an extra $40,000 in excess hospital charges, and a ten percent excess risk of attributable mortality.

So the argument is that although

these events are uncommon, they're very serious events that have profound implications for the patient care and for the healthcare system. So I think that addresses the issue of importance, preventability, and the coding issue.

MEMBER HALPERN: What happens to those patients who are like trauma or, in my case, ruptured aneurisms who wind up with abdominal compartment who get opened at the bedside but not for dehiscence reasons but because you need to relieve the inter-abdominal pressure?

MR. ROMANO: This code is specific for a situation where the wound is disrupted, where the wound opens up spontaneously after surgery and the patient has to go back to the operating room to have that defect fixed, so it would not apply to a compartment syndrome kind of situation, because in that case there may be some leakage, but there is not a spontaneous fascial dehiscence that's

occurring.

MEMBER CIMA: Just to respond to Dr. Romano's comments about the studies that showed the mortality and the assessment, yes, there is an association. All the literature supports that.

The problem with those studies is that you, and as noted in the original study in '89, this is -- and also in the pediatric studies -- a very high percentage of these are associated with some other process, usually an intra-abdominal infection.

Therefore, and this is one of the big issues we see in the administrative data set tying to economic factors, is, yes, wound disruption and requiring a closure does add to hospital length of stay and cost.

However, the vast majority of the time in the literature and in those studies that were cited, there was a secondary or probably the primary cause of the disruption was an intra-abdominal infection or intra-

abdominal catastrophe which then is the primary reason why the patient has increased length of stay and cost, and the secondary coding of the dehiscence was partly associated with it.

It's a coordinated confounder that goes with it. They move together, and to assume that this is the best marker is probably a false statement.

MEMBER STAFFORD: I would agree with that. I think what you're saying is that this is really a surrogate for something else that's going on, and should we -- should a measure be intra-abdominal, deep intra-abdominal infection associated with a procedure? That might get you more bang for the buck.

From a surgeon's standpoint, that's really -- when we stand up at M&M and say we had a wound dehiscence, the first question everybody says is, "Well, what caused it?" That's what they want to know. It's not

that there was a dehiscence, but what was the underlying cause for that?

So I think in that sense it's important, and then the other question I would have for Dr. Romano is since this measure has been endorsed, do you have any evidence that it's made any difference in patient care and outcomes?

CO-CHAIR MORRIS: That's an important question. Dr. Romano, do you have a response?

MR. ROMANO: Yes. So the -- I can pull up the national numbers. I know that some individuals around this table have been involved in local efforts to address this problem, and so I can't speak to those efforts.

I will say that the other fact -- the risk adjustment model for this indicator does include heart failure, hypertension, chronic lung disease, obesity, anorexia or weight loss, and alcohol-related conditions,

so many of the conditions that are associated with poor tissue are included and have the expected effects in the risk adjustment model, as well as, of course, age and some underlying abdominopelvic conditions.

Also, to address another question that was raised, there was a linkage study that was done with NSQIP. It was reported. It is one of the studies that was cited in the submission. It was an HSR a few years ago.

In the NSQIP linkage study, we did find that about 35 percent of the events were missed, so there was some proportion of the cases that were missed based on linkage with NSQIP.

Now, admittedly, the definitions are a little bit different, so we couldn't drill down to the finest level in terms of whether those cases that were missed in the PSI but were picked up by NSQIP, what was the nature of those cases. We can't say.

Finally, again I refer you to the

paper from the VA that's in front of you that comments on some of the associated conditions. Forty-three percent of the events were associated with the evisceration of abdominal contents.

Twenty-eight percent had fluid draining from the wound. Thirteen percent had some coughing or physical exertion that may have triggered the dehiscence. Wound infection was only found in nine percent, and only three patients with a perforated hollow viscus.

So at least in this series it was a relatively small number that had a wound-related infection, although certainly I concede the point that this is to some extent a marker of more serious things that may be going on.

MR. BOTT: If I can jump in for one second, this is John Bott with AHRQ. So this is a question we've been asked in past steering committees. Have we seen an

improvement as a result of the use of the measure?

Our budget and our focus for AHRQ for the quality indicator project is developing and maintaining the quality indicators and improving them where there is opportunity. So that's what our focus is, and so when users have questions about the measures, being responsive to that.

If they have suggestions for improvements or how they could be refined, we catalog that and prioritize it and act on those. Our focus has not been the cataloging and tracking so much of how people have used them and where but to be responsive to what their concerns are to continue to improve the measures.

We did complete, of course, the application, the NQF form, and it does ask who is using them, and we put in there who we were aware of that we are readily aware of by people contacting us, but we have not

canvassed the many groups who use the measures to understand the way in which they've made an impact now.

That's not to say that couldn't become an aspect of AHRQ's future work in supporting the indicators, but that, of course, is resources we currently do not have in the quality indicator projected.

MR. ROMANO: I did pull up the national trend data, which shows a decrease from 3.2 per thousand cases in 1994. At the advent of the quality indicator program, it was 2.62, and the most recent data are 2.48, so it suggests a modest downward trend, but I would, of course, be reluctant to attribute that to the use of this indicator.

MEMBER MORTON: I have one comment is that I agree with Bob. This is probably a surrogate measure. It's looking at other processes of care that impact the outcome of dehiscence.

I liken it a lot to someone having

a temperature. We're not sure exactly what's going on, what's the reasoning behind the temperature, but it allows you to kind of dig in a little deeper and see what's going on.

Utility for these PSIs are clearly around the fact that they're readily available. You search where there is light, and all of these different measures are actually fairly easy to obtain, and correct me if I'm wrong. I think they've been endorsed by a lot of the state public reporting agencies around quality.

This one is probably maybe not the strongest of the bunch for the PSIs, but I think there is some utility to having these things available. There are a lot of hospitals that don't have NSQIP, that don't have registries, but they can get this sort of data through the PSIs pretty readily available through the billing information, and then you can kind of dig in and see what's going on.

MEMBER DILLON: Yes, but that

should be the institutional's M&M process, rather than at the sort of the national level. I guess one thing I would say is that I think it's very important that if we're going to support a metric that it's a meaningful metric, and obviously that presents a challenge here.

I will propose, perhaps, an idea, at least on the pediatric one. Since there is so little data to support it, one thing that I would ask that we could consider if the NQF is amenable to it would be to have a HRQ, NQRI, and NSQIP, a peds NSQIP process get-together, and take a look at this as a true outcome measure in terms of quality.

The peds NSQIP will start having data in about six more months, so there is the possibility that we could combine forces before considering this as a metric that should be passed. Just a -- just a consideration.

CO-CHAIR MORRIS: Any other

comments or thoughts?

MEMBER STAFFORD: Just two other comments, just thinking about it and thinking about it more as, I think, what most of us surgeons think about this as being a surrogate marker. I really liken this to the IHI trigger tool where you go through a chart and you, you know, you see --

You're trying to find out about hypoglycemia, so you actually look through the pharmacy records for how many amps of D-50 were given during a period of time, and then you go through the chart and you look to see what the reason for that was. I really think that's how I see this, and I think that's how most of us use it at M&M on a daily basis.

Now, granted, folks in private practice, depending on where they are, may not have a robust M&M, and so I think we also have to remember that those of us who are in academics there is a whole other world out there, and how they use these indicators may

or may not be valuable to them, so I think for a lot of us we may have a bit of that bias.

All of that being said, because I think it's clearly something to look at, but whether it should be an indicator that gets an official stamp which then CMS and other folks may use as a pay-for-performance issue is a whole other target.

Again, that's what I kind of want to keep us all thinking about is when we put a stamp on something from the NQF, that has some say in the real world, and these things then tend to get picked up by other groups, Leapfrog and CMS, et cetera, so I really think it's something to think about.

MEMBER CIMA: That's the big issue here, the difference between trigger tools and IHIs. It's internally done and developed. Using this as a safety indicator, is it truly a marker of safety?

That's a very different thing, you know, and I would submit that there hasn't

been any significant association with this marker, because it is so infrequent that hospitals rarely ever do anything with it.

I can show you in our -- you know, looking over 100,000 operations a year, this is -- this is so low a frequency event that we can't do anything with it, so it just is there, and so it's not meaningful.

CO-CHAIR MORRIS: Any other comments? All right. So basically, to sum up, we talked about how this is a patient safety measure, so therefore it should be a rare event, that it's really more focused on impact on the patient, rather than prevalence, and that seems appropriate, but on the other hand, that it doesn't seem to really be preventable or actionable for the most part.

Richard brought up the excellent point that central line infections were not deemed preventable or actionable previously, but still I think that that's the overlying opinion of the group that these aren't really,

for the most part, preventable or actionable. So the implications for patient safety may be pretty modest, but the implications for provider -- for potentially negative provider sequelae may not be quite as modest.

I think that something else that came out of this discussion that we could potentially recommend to NQF is that measure developers who would like for measures to be maintained provide support for the impact of the measure since the time that it was previously endorsed, so basically providing a little more data about the actual impact.

I think that we should consider charging the developers with this. It seems like the appropriate place to put that responsibility, and then the developers can figure out best how to do that within their resources.

Let's -- if there is anything else, if anybody feels like their point was not summarized here, please speak up, and then

we'll move on to the vote.

MEMBER CIMA: I will say one other thing for the developers. If you're going to cite literature, it's very important to cite it correctly.

That really turned the argument on this was, you know, the level of -- we're reading this. I know it's a lot to put it through, but if you want us to read it, then it's got to be accurate, and this was blatantly not accurate.

MR. ROMANO: I would like to apologize for the misleading presentation of that evidence in the submission. It was not an intentional effort to mislead the panel, and, again, I tried to provide the best and most current papers here for your review, so I apologize for that.

CO-CHAIR MORRIS: Okay, let's go ahead and move on to the vote, then. Does the measure meet NQF criteria for importance to measure and report? Just so everybody is

aware, this is the pediatric measure, 0367. Five said yes. Fourteen said no.

Now, we need to provide a reason for maintenance measures that we consider -- that we deem not important, correct? Okay, so we'll stop here, since we're not really talking about inactive or emeritus or Hall of Fame status.

So we'll move on to voting on 0368, and this is the adult measure, unless anybody has anything else to add. Okay.

MR. ROMANO: Can I ask a question? Just it would -- it would be helpful, because this is a contradiction of a previous NQF endorsement panel, so I think it will be important for the record to state clearly why it's no longer considered important.

CO-CHAIR MORRIS: We had our discussion. This will be transcribed and posted. Is there -- is there more that we should say about that? I mean, just basically I'd be re-listing the discussion that we just

had.

MS. MURPHY: We will have all of the information written up in the context of the measure regarding the discussion.

DR. BURSTIN: I suspect it would probably be grounded in 1c, around the outcome or the evidence in terms of the evidence for the measure focus being in question in terms of the link of the measure to the outcome of interest, but, again, we can clean that up a bit.

CO-CHAIR MORRIS: Okay. Let's go ahead and vote on the adult measure. Are you ready, Jessica? Okay. So we're voting on the adult measure now, importance to measure and report.

Okay, we're going to do a show of hands. This is the last vote for today. Oh, here we go. Okay, six votes for yes, 13 votes for no.

Let's open the phone for public and member comment regarding this last

portion. Anybody on the telephone?

Okay, the last section we are going to be talking about at 1:30, and Dr. Torchiana is going to take over here. We just want to make sure all of our developers are available.

CO-CHAIR TORCHIANA: So the 1:30 topic is under Attachment 6, related and competing pediatric and congenital cardiac surgery measures. So when Arden and I were talking earlier today, we hoped that we would get to this topic before 1:30, and the question is do we have our developers here.

There's Dr. Jacobs from STS is here. Kathy Jenkins from Children's Hospital of Boston? On the telephone? Kathy, are you on the telephone? Not as yet, and then there is an AHRQ developer, Patrick.

So we have a choice of waiting four minutes, or we could have the Children's developer go last, assuming she comes on by 1:30. Why don't we do that, rather than

flipping the agenda around?

MS. MURPHY: Well, I wonder, only from the standpoint of everybody hearing all of parts of the discussion, if maybe we can do just a couple minutes on this gaps document. That's the very last item on the agenda. Would that be okay?

CO-CHAIR TORCHIANA: Well, that's what we had discussed doing. My thought was if we ask the STS then AHRQ to basically describe the rationale behind their measures that hopefully by that time we would have filibustered up to 1:30, and we would have our Children's representative on the phone.

Is it possible she's on the phone and muted?

DR. BURSTIN: Operator, can you please check if Dr. Kathy Jenkins is perhaps on the non-speaking line?

OPERATOR: She is not online at this time.

CO-CHAIR TORCHIANA: She is not.

Okay, I'll fold my tent. Let's do the other.

MS. MURPHY: So, in the materials that you received for this meeting, and I'm referring to the document that had the nine attachments, there is Attachment Number 8. I don't know if we can -- Alexis, are you going to be able -- okay. Good. And all this is just a little bit of an introduction to this. That's all I want to do to have you take a look.

One of the items that we are doing with each of the steering committees as they look at their area of assignment is asking them also to look at it in terms of where there might be gaps in measurement.

So what this document provides to you, down the left side are a list of the surgery measures that are currently NQF-endorsed, and the yellow highlight focuses on those measures that are in consideration in the work you're doing right now.

So you can see the ones you're

working with right now and the remainder of the NQF-endorsed surgery-related measures, and then across the top of it what we've done is identify different domains of care or service within which you might consider whether or not there should be a measure.

All that we are asking that you do, and not do it today, but to take a look over this and identify where you believe either in the current topic areas, or you can add additional topic areas across these domains, where we should look for additional measures at some point in time.

CO-CHAIR TORCHIANA: Melinda, could I ask where the domains were defined?

MS. MURPHY: If you're looking -- are you looking at where the definitions appear on here or --

CO-CHAIR TORCHIANA: No, I'm just curious that so many outcomes measures are under care coordination and management, since there is no outcomes domain.

MS. MURPHY: In the discussion of work done with the National Priorities Partnership and some of the work that's been done with the new group looking at measurement, they have identified and defined -- to some extent helped us in identifying where topic areas fit within the domains. Helen, is there anything?

DR. BURSTIN: Most of these correspond to the domains that are recently released, National Quality Strategy, so this does relate to that. I'm not sure patient outcomes are there, per se, but those relate quite well to the domains that the Secretary recently promulgated.

MS. MURPHY: So this is a homework assignment. So if you would take that away, take a look at it when you have an opportunity and provide us some feedback.

DR. BURSTIN: Peter, I wonder if you want to mention the pediatric surgery discussion you and I had recently.

MEMBER DILLON: I can. It's still embryologic, but we will convene or will discuss with NQF and take these surgical measures and focus them on the 18-and-under population with a group of multi-specialty surgeons, and we'll take a look at ones that we think should be then brought forward as potential measures within the children's surgical care. We've just -- we've been trading, you know, the NQF lists, and we'll start to focus on those.

DR. BURSTIN: So far, the idea there was to see how many of the measures that are labeled as adult are labeled as adult just because that's what the developer did and wasn't really thoughtful and could consider whether, in fact, if you looked, some of those measures could be applicable to children, as well, with some adaptations. So I just wanted to let you guys know that's happening.

MEMBER DUTTON: Well, there's a -- well, a lot of these are -- many of these have

anesthesia implications. There are no specific anesthesia or pain management measures in here, so I need to think about it a little bit, but you'll get some from us, as well.

MR. SILBER: Hi. Jeff Silber on the line. Hello?

MS. MURPHY: Yes, Dr. Silber?

MR. SILBER: Hi.

MS. MURPHY: Hi. So we're just finishing up one other --

MR. SILBER: Okay.

MS. MURPHY: -- quick discussion. So that's the -- that's the gaps work for you, please. Thank you very much, and we have not identified a deadline, but we'd appreciate it whenever you can do that for us. If you just send it back to either me or Alexis or Jessica, then we'll compile it all and give it back to the Committee as a whole.

MEMBER ROGERS: So, Melinda, just to make sure -- this is Terry. Hello? This

is consistent with my past academic experience. I just want to make sure about the assignment. I see --

MS. MURPHY: I bet you always made As.

MEMBER ROGERS: I'll just let that one go. So we're to review again the initiatives that we've looked at for the past day and a half and see whether in these new gap areas there are some issues that we may have missed and that we should want to consider. In those, we mark a box plus, minus, a comment. Is that -- is that what you're asking?

MS. MURPHY: Right, except I think it's a little more simple in terms of we're not asking that you go back and take a look at the specifications of those measures, the topic areas. So are there any topic areas that we should now be looking for measures to fill gaps? Thank you.

CO-CHAIR TORCHIANA: So, Melinda,

would you like us to start with the pediatric measures, or should we go to the failure-to-rescue measures?

MS. MURPHY: I guess this is my moment of confusion, one of many. We had the discussion about the failure to rescue measures, so discussion about them will be -- the next discussion about the failure-to-rescue measures relate to related and competing, which we are not taking up today.

The only ones that we're taking up today in terms of related and competing are the peds measures in terms of the ones that are related and competing, so that is measure PCS 1809, which is STS Measure 2109, which is Children's Hospital Boston, and 0339. This is the mortality measures, which is AHRQ measure, and then the related volume measures.

MR. SILBER: Could I ask a question? A little bit of confusion, because I have a bit of a window. We had scheduled an hour.

I thought that I was calling back in today about my FTR measures because of an issue of distinguishing them between the original, which is my measures, and the AHRQ measure. It's not going to be discussed today, because otherwise I don't know if you need me on the call.

CO-CHAIR TORCHIANA: I think we won't need you on the call today, Dr. Silber.

MR. SILBER: Okay. Is this because it's being rescheduled or because it's already endorsed and it isn't an issue?

MS. MURPHY: Dr. Silber, the discussion about them with respect to related and competing will be scheduled with this Committee at a later point. It will be a conference call, and we will notify all of the developers and invite and be certain we have them available for that discussion.

MR. SILBER: Okay.

MS. MURPHY: Thank you.

MR. SILBER: Can you hold for just

one minute? So I guess it's just not today. Okay. We will -- we didn't understand, and sorry to bother you, then. Thank you.

CO-CHAIR TORCHIANA: Thank you. Okay. So the next section is Attachment 5, related and competing measures. We have three developers. Melinda has outlined the three measures that we're here to discuss.

So we'll begin with the developers. Let's start with STS, and I would ask just to sort of establish the format of what we hope to accomplish in the next hour that we'd like to begin with each of the developers describing their measures. We would ask that you refrain from any comments on the competing measures during that preliminary step.

We then would ask that the Committee discuss the various measures and discuss the issues raised with the goal that as we get through this conversation we will be able to help define issues that we might ask

the developers to return with around data evidence or modifications so that this can be brought to closure in a subsequent conference on the telephone. So let's start with the developers, Dr. Jacobs, for the STS measures.

MR. JACOBS: Thank you. My name is Jeff Jacobs. I'm the Chair of the STS Congenital Heart Surgery Database, and the background for this discussion is based on the fact that in pediatric heart surgery there is a vast array of operations, each one of those operations being done in limited amounts.

So, unlike adult heart surgery, where you have hundreds of thousands of coronary bypass grafting operations or aortic valve replacement or mitral valve replacement, in pediatric heart surgery there is over a hundred basic types of operations that are done on the hearts of children, and the risk ranges tremendously from one operation to the next, and there's not high volumes of any of those operations.

So, in order to assess the mortality after congenital heart surgical operations, a methodology is needed to assess the complexity of the operation performed, and the measure that we propose is based on a tool called the STS/EACTS Congenital Heart Surgery Mortality Categories.

What we did was we had -- we pulled data from the STS database and our counterpart database in Europe, the EACTS database, that was gathered using identical nomenclature and terminology, and based on that pool of data we had access to over 75,000 pediatric heart surgery operations that were divided into 148 major operative procedures.

Those were then analyzed and grouped into five categories or buckets of increasing operative complexity set up in a way to maximize the ability to discriminate between one category and the next category and to maximize the similarities within any given category.

What we then had was a tool that had five categories of complexity with Complexity 1 being operations that are the least likely to result in mortality and Complexity 5 operations that are most likely to result in mortality.

In Category 1, mortality is around 0.5 to 1.0 percent, and in Category 5 mortality is over 20 percent. By having these five categories, one can then report mortality using similar categories that allow for meaningful comparison between institutions.

STS feels it is very important to use a tool that functions in a clinical database, rather than assessing outcomes after pediatric heart surgery from administrative data because of multiple publications that document the flaws and coding of congenital heart lesions utilizing ICD-9 codes and administrative data.

The one example I would give for that is the most talked about and focused upon

operation in pediatric heart surgery that's used to assess programmatic performance is the Norwood operation. This is an operation that has between 15 and a 25 percent mortality after the operation, so one in five to one in four babies that have this operation do not survive and die before they go home.

In a clinical database like the STS database or other clinical databases, there is a procedure called the Norwood operation, so you can quote for it. If one wants to find the Norwood operation from an administrative database, one would have to create an amalgamation of inclusionary and exclusionary criteria based on 15 different ICD-9 codes, some of which are required, some of which you cannot have, some of which you can have.

It's amazing to me that based on this amalgamation of 15 different codes one can even identify this operation. That's one example of the weakness of administrative

coding.

So I think what we've tried to do through STS is create a system to stratify operative complexity that can be used both in clinical and administrative databases, but we emphasize that we feel that it should be done with administrative data -- I'm sorry, with clinical data because of the flaws of coding pediatric heart surgical operations with administrative data.

I think -- I don't want to take any more time right now. I think that's a pretty rapid summary of the way our method is -- methodology works and how we apply it.

CO-CHAIR TORCHIANA: Thank you, Dr. Jacobs. Do we have Dr. Jenkins on the phone now?

DR. JENKINS: Yes, this is Dr. Jenkins.

CO-CHAIR TORCHIANA: Could you speak up a little bit?

DR. JENKINS: Yes, I'm here. I'm

on the call.

CO-CHAIR TORCHIANA: Great. Could you describe your measure, the Children's Hospital measure?

DR. JENKINS: Sure. The measure that we've proposed in this entire process is known as RACS(1). It was developed a number of years ago, and it's been widely used recently, over the last eight years or so.

It's probably the most commonly used measure of mortality in the United States and across the globe. What we've heard through this entire process is a lot of sessions about what's preferable about our methodology over some of the other methodologies that have been put forward, and let me just speak to that specifically.

I think it's important to say that we've put forward our methodology as it was originally developed and validated years ago as a methodology that can be used in administrative data, which has been -- there's

been a lot of conversation about that use, as well as non-administrative data, and I think that's important for the Committee to know.

In fact, one of the reasons that we went forward with our pure RACS methodology, even though there was already the AHRQ methodology that had been approved previously, was the fact that the prior AHRQ methodology was put forward in a slightly different version than the way our measures had previously been developed and validated in terms of the additional risk factors that were incorporated for some reason that I believe related to some internal AHRQ harmonization but was different than the way the measure had been done, and also because the original AHRQ methodology only specified the use of the methodology in administrative data and did not incorporate the use of the methodology in non-administrative data.

So what we've put forward is a measure that includes risk categories for

procedural complexity, as well as additional risk factors wrapped up together in a multi-variate model that yields an SMR for institutional performance and covers, in general, roughly 85 to 92, 93 percent of an entire pediatric case mix across most case mixes.

In terms of our -- and I will say something in response to the specific issue that Jeff Jacobs brought up about weaknesses of administrative data based on ICD-9 codes. While it's definitely true that there are not good ICD-9 codes to detect surgical procedure for Norwood Stage 1, there is an excellent ICD-9 code for the diagnosis, which is HLHS.

So the algorithms that have been built to detect those specific procedures have in general worked out pretty well, and every data set that I've ever looked at using RACS(1), the category for the Norwood type procedures, which is the highest risk category, has always had a mortality rate that

was distinctly higher than and distinctly different than any other of the other categories.

So I don't think there's a lot of problems with identifying those procedures, at least in the sense that the category is finding procedures of especially high mortality.

I would like to say something about weaknesses with the five category mortality rates that I believe has also been proposed as part of this conversation, just so the Committee is aware of them.

First of all, when we talk about five-category mortality, whether we talk about it by RACS(1) or the new STS categories or a prior version of Aristotle categories, the categories in general work well for the procedural complexity.

The problem is that you take a case mix which is typically around 200 or 250, maybe 300 cases, and you divide it up into

five or six categories, and, as everyone knows, what you've done then is you've essentially distorted and diluted any statistical power to make any meaningful comparison, because the confidence limits around the five-category mortality rates are really rather -- are very large, even in large institutions. So almost by definition from statistical power alone you've diluted your ability to find meaningful differences.

In addition, not surprisingly, those categories then don't account for other clinical risk factors like the age of the patient and whether they have other anomalies and whether the infant is premature, and so the categories alone really don't account for all of the variation.

In general, the categories will give an area under the RSC curve. Some were in the range of .7. You need the additional risk factors to bump it up beyond that.

So it's a combination of the small

sample sizes and typical pediatric case mixes and the fact that there are procedural complexity variables beyond procedure that make it much better to wrap up the entire measure into an SMR that brings all of that together and adds statistical power.

CO-CHAIR TORCHIANA: Okay. Dr. Jenkins, could you wrap up your description?

DR. JENKINS: That was the end of my comments.

CO-CHAIR TORCHIANA: Thank you so much. Who will speak for AHRQ? Patrick?

MR. ROMANO: I will, yes. This is Patrick Romano again. So this is easy for me, because I think we're very clear and direct that this AHRQ indicator is based on the work of Kathy Jenkins and her colleagues at Children's Hospital Boston, and so we defer to them regarding the specific issues in the application of the RACS methodology. It's based on the same RACS methodology.

So the different -- of course, you

know, in terms of the process by which this became an AHRQ quality indicator, fundamentally AHRQ and its contractors troll the field looking for quality indicators that can be applied to administrative data sets.

Kathy Jenkins' work, of course, came to the attention of AHRQ. It's in the public domain, and we actually had a number of conversations with Dr. Jenkins and her colleagues about the specific operationalization of the indicator.

So there are some differences which are really technical differences that relate to how the indicator is presented. Kathy described the construction of an overall SMR, a standardized mortality ratio, kind of an observed-to-expected ratio.

The AHRQ quality indicator software spits out a risk-adjusted mortality rate, which is simply an SMR multiplied by some overall average mortality rate, so it converts it into a percentage.

The other differences relate to the risk adjustment. Those could easily be reconciled. Fundamentally, I think both approaches adjust for the RACS categories that Kathy has mentioned.

They adjust for age and birth weight or prematurity. The adjusters in the AHRQ are slightly different, because they come out of a certain risk adjustment structure that's embedded into all of the AHRQ quality indicators.

For example, there is a specific indicator for patients who are transferred in in the AHRQ model to account for the fact that some of those patients may be higher risk in ways that aren't directly captured elsewhere in the risk adjustment. So those minor differences could be -- could be reconciled.

The final point that I want to make just to the general issue is that there have been -- there's been quite a bit of experience, I think, with this measure, both

with Dr. Jenkins' team and others.

In fact, there was a comparative study, O'Brien and Clark, that was a head-to-head trial, head-to-head study that basically showed that these two approaches, the STS-based approach, the RACS-based approach, are functionally equivalent in terms of discrimination, in terms of the ability to discriminate between kids who die and kids who survive the hospitalizaton.

So from the statistical standpoint it's very hard to discern a meaningful difference in the performance. The fundamental difference is that one indicator is based on a registry system, which is very important for quality improvement, and the Children's Hospital Boston and our measure, the AHRQ measure, are based on administrative data.

CO-CHAIR TORCHIANA: Comments from the Committee?

MEMBER HALPERN: Do you have a --

since we didn't get to review your measure, it's clearer in the Children's Hospital measure that they are looking at the low-risk patients, it seems to me, because they're saying RACS(1) risk category, but it's not clear in yours how you discriminate between severity of disease.

MR. ROMANO: It's the same, actually. There's a -- the RACS is a five-level scoring system, and so the RACS categories are put into a multi-variable risk adjustment model, so each of the categories carries a certain additional risk that's associated with it. I believe, Dr. Jenkins, I believe -- is that the same approach that you used?

DR. JENKINS: Absolutely, yes.

MEMBER HALPERN: So, do you report it by category?

MR. ROMANO: The AHRQ measure reports it as an overall composite in the same way. It's just the only difference is that

the Children's Hospital Boston reports it as a ratio of observed to expected, and the AHRQ software translates that into a risk-adjusted mortality rate.

Users can, of course, drill down and look at the stratum-specific mortality rates for each of the five RACS levels, but, of course, that's more difficult for consumers, purchasers, payers, other stakeholders to understand, so the preferred measure from the standpoint of AHRQ's stakeholders is the composite measure.

MEMBER DILLON: What are the metrics for your composite measure? What are the units?

MR. ROMANO: It's a risk-adjusted mortality percentage.

MEMBER DILLON: So it's -- I'm sorry. That's what I'm getting at, just a straight percentage.

MR. ROMANO: Yes. It's derived by the ratio of the observed to expected

multipled by the overall mean average, so it's the standard approach for what we call indirectly standardized mortality rates.

CO-CHAIR TORCHIANA: Peter, I think it's a composite in that it's a composite of all risk levels.

MEMBER DILLON: Right, because the STS results in five. You've got a score for each of the five levels.

MR. JACOBS: And the rationale for that is that imagine you're a mother with a child that has the highest level complexity patient. You might want to be able to find out how a given center performs caring for patients of that level of complexity, and that might not be possible if you just get an overall aggregate score.

So it was felt that it's important for patients and patients' families to be able to access how a given institution performs both in the low levels of complexity and in the high levels of complexity.

MEMBER HALPERN: I guess you feel that your model, your risk adjustment model accounts for a case mix by its risk adjustment.

MR. ROMANO: Well, it does account for case mix. I can't argue with Dr. Jacobs' point. It's just that different stakeholders have different needs, and so some might prefer to see a single measure that's specific to a certain risk level.

Others will prefer to see a composite measure that incorporates both low- and high-risk patients and adjusts for the difference, but, again, statistical performance is very comparable.

MEMBER HALPERN: So it seems to me that yours and the Children's Hospital are very similar, just a different way of reporting the same data, whereas the STS has a distinct purpose.

So, to my mind, it's really comparing the AHRQ and the Children's

Hospital, because it seems to me that the STS database is reporting it in a different way. They're reporting by category so people can actually see by category, by risk, you know, by the severity of disease.

MR. ROMANO: Right, and, in fact, from what I understand, there is a strategy both to deal with competing measures and complementary measures, and it might be viewed that while the AHRQ and the RACS are competing measures, the STS measure could be viewed as a complementary measure, because it's taking a fundamentally different approach to provide access to a different type of information.

CO-CHAIR TORCHIANA: I think that's potentially a way forward, but I guess I'd say this is very reminiscent of a long conversation we had yesterday, with apologies to the developers who weren't in that conversation.

This is quite similar to the aortic aneurism repair conversation in that

there is a sort of tension between adding things together and getting more ability to discriminate at the institutional level versus splitting them up into more appropriate, more homogeneous groups that thereby then lose some ability to discriminate. I do think it's not unreasonable to think of that as complementary rather than competing.

Could I ask Dr. Jacobs the plans or current status of the reporting this to the public? If this is for families, is this report to the public currently by the participating STS institutions, or is that a future plan?

MR. JACOBS: So, as some people in this room probably are aware, the STS has moved forward with public reporting of cardiac surgical outcomes fairly aggressively in the past year, and the approach we took I can speak to in quite a large amount of detail, because I chair the Public Reporting Task Force for the Society of Thoracic Surgeons.

What we decided to do was initially to report from the adult cardiac surgery database CABG outcomes, and we've been doing that for just about a year now. We wanted to work out the kinks of public reporting of cardiac surgical outcomes using the CABG outcomes first, but that's really a platform that will then be expanded in adult cardiac, also to aortic valve and mitral valve, and our other databases will be expanded to both thoracic and congenital.

So our intent is to do just like we do with the CABG outcomes and to publicly report these outcomes, as well, using similar strategies, and I think that's something that's going to happen quite soon.

MEMBER STAFFORD: At what level are you reporting those outcomes, provider level, facility level, group level?

MR. JACOBS: So the CABG outcomes are reported both at group level and facility level, so one can go to the STS website and

see the CABG outcomes for a given hospital or a given surgical group.

Up until now, the decision has been made not to report it based on an individual provider based on the concept that outcomes after heart surgery are dependent on so many elements of the team and not just the surgeon independently. One could say heart surgery is a team sport, not an individual sport, and based on that so far we've reported it at the group level and at the hospital level.

MEMBER DUTTON: Can I ask a question about the data? Oh, I'm sorry, Terry. Go ahead. Dr. Jacobs, in the data you've gathered, what does it show about breaking the cases down into five different categories? Do you find centers that are good at simple operations and bad at hard ones --

MR. JACOBS: Absolutely.

MEMBER DUTTON: -- and vice versa?

MR. JACOBS: You find several

things. First of all, you find that this methodology has the ability to identify outliers, even though you've broken it into five separate strata or categories. So we don't need to pool all the data together to identify outliers. Outliers can clearly be identified in each of the five categories.

Second of all, we've found that there are centers that don't perform surgery at all in the most complex category, so there are centers that don't even do Category 5 surgery, but instead they refer the patients elsewhere.

If one pools all the data together to create one score, then when one looks at that score, it's impossible to know whether or not that center is performing the high-complexity surgery. So, you know, not only can you identify outliers within each category, but you can also with this methodology identify centers that don't perform the high-complexity surgery at all.

MEMBER ROGERS: I am speaking, hopefully, for the Committee to be reminded what our specific responsibility is today and what the impact of that might be. That is, if we were asked to make a decision, do one or two of these go away, or what happens?

CO-CHAIR TORCHIANA: I don't think we're being asked to make a decision today.

MEMBER ROGERS: Okay.

CO-CHAIR TORCHIANA: We talked about that at lunch amongst this end of the table. The goal is to try to have a fruitful discussion to give feedback to the developers and I think to identify modifications or additional information that might help lead us to a future decision, but the decision is not to be made today.

MEMBER ROGERS: Thank you. Then I would like to make one comment. Having played in both the arena of administrative data and clinically derived data, recognizing the complexity of this specific issue and also

paying homage to the seriousness and commitment of all players involved in all three of these, I would have to err on the side of going with a clinical database for this kind of process.

To the extent that that kind of conversation can take place between and amongst, that's great. If not, then I would bend on the side of the clinical database on an ongoing quality measurement process.

MEMBER MORTON: We heard yesterday the penetrance of STS for adult hospitals was pretty high. It was like 95 percent. I'm curious about the penetrance of STS for pediatric cardiac hospitals.

MR. JACOBS: Right. That's an excellent question. STS has done several manpower surveys, and based on those manpower surveys we estimate that there's 122 hospitals in the United States that do pediatric congenital heart surgery, and that estimate has been validated through other sources of

data, as well.

Right now, STS receives data from 98 of those 122, and if we look at the 20 largest by volume, we get them all. If we look at the 20 that are listed in the U.S. News & World Report as the centers of excellence, we get data from 19 of those, and 20th I think is going to start sending this year.

So it's -- I think the penetrance of the STS database is fairly high. The ones that we're missing are low-volume programs, and we're making every effort to encourage those low-volume programs to participate, as well.

MR. ROMANO: Could I just ask -- I'm sorry. What percentage have agreed to public reporting of those data that are being contributed?

MR. JACOBS: That's a good question. So right now we haven't even asked the congenital heart surgery database yet,

because that initiative hasn't started. In the adult heart surgery database, last year was the first year that we did public reporting.

As most people know, it was done on a voluntary basis, and in year one we had 20 percent of the programs participate and publicly report their data. Now, in year two, that number is up to 39 percent right now, so going from one year to the second year we've already doubled the participation, essentially, and I think that it will continue to climb.

As long as it's voluntary, I doubt it's every going to get to be 100, but the fact that we doubled it from year one to year two tells me that more and more places will do it, and if enough people do it, it becomes functionally mandatory.

DR. JENKINS: This is Kathy Jenkins. I'd just like to make one more comment so people are aware that in response

to Patrick Romano's earlier comment about the ease of harmonizing our methodology with the AHRQ methodology, I had proposed that very early in this process, and through the NQF rules we had been told that we weren't really allowed to bring things together at this stage, so --

DR. BURSTIN: I think that actually only had to do with -- this is Helen, Kathy -- because this measure was in a different project, but I think now that we've actually sort of brought them somewhat together, I think those options are completely on the table.

DR. JENKINS: And I would just like to reiterate our willingness to do that.

MR. JACOBS: And I think from STS' point of view it seems to me that it would be, first of all, very nice to see the AHRQ and the Boston Children's measure harmonized.

I think that ultimately STS would view that as a favorable move and hopefully

would also -- STS would also be supportive of seeing a harmonized AHRQ RACS measure and the STS measure ultimately both being endorsed as complementary measures, because I think from the discussion that we've heard today each measure brings some very nice features to the table. Then we could have two complementary measures, a harmonized measure of AHRQ and RACS and a complementary STS measure, and that's something STS database would be very supportive of.

DR. JENKINS: And that's why I just went through that conversation, Jeff, though, because, as I said before, our methodology was originally validated in a registry which is not the STS registry, but the Pediatric Cardiac Care Consortium registry is very similar, and so I don't see any reason why an SMR using the RACS methodology even in a clinical registry with STS is not also a possible measure that could be useful publicly.

MR. JACOBS: That's exactly what I said. I think that I would be very supportive of that happening, and I think it could be viewed as a complementary measure to what STS is proposing, so I think both of them together are additive, rather than competitive.

CO-CHAIR TORCHIANA: Could I ask for input from the Committee on that question, the notion of having complementary registry and administrative data coexisting?

MEMBER HALPERN: I think that's a -- I think that they both bring separate things together that are equally important.

MEMBER DILLON: The problem that you have to be careful of is that we don't get into -- we don't get competing or, I was going to say misleading, but confusing information, and I'll cite the difference between UHC and NSQIP type data right now, which are 180 degrees opposite.

So the last thing you want is some mother looking up, "Well, wait a minute. This

number says good, and this number says not so good. What do I do?"

So the question I have is if you rolled up all your STS data, what's the -- has anybody looked at the correlation between a unified measure of your overall STS results and how it might correlate with the administrative studies, the RACS study or with the AHRQ?

MR. JACOBS: Well, I think that all three of these methodologies can be applied to administrative data for registries, and the reason that I said that they were complementary was not because administrative data and clinical data is complementary but because what the STS does with the STS categories is somewhat different from what the RACS and AHRQ methodology do.

I think there is value to having both pieces of information. There is value to having the individual outcomes of the five categories at a given program, and there is

value to having the overall composite, and that's why I think they're both complementary. It's not a complementary nature of clinical or administrative data, because the clinical and administrative databases can function with any of these three tools.

We have not rolled all five categories into one score, because we think there is value in having each of the five -- the information of each of the five categories the way we present that. However, that's something that could certainly be done if it was felt that that would be a beneficial part of the process, as well.

DR. JENKINS: And this is Kathy Jenkins. I understand the point of the question about confusing to the public, and I've never really specifically seen that exact academically done heads-up comparison that you're actually looking for, but having seen many, many, many reports in various ways using both clinical databases as well as the

administrative databases, I've never seen an especially confusing signal with that sort of one looks up and one looks down kind of a problem that you're alluding to.

I don't think the problem here is really with the administrative versus clinical signal. We're really just talking about how to get the risk adjustment right so that it really adjusts for complexity across such a diverse case mix in a way where we can make sense of the information and have enough statistical confidence that there is clarity.

MR. ROMANO: And if I could address that point, so AHRQ, I think we defer to Dr. Jenkins in terms of her methodologic expertise in developing this measure, but let me just explain how the complementarity works to some extent.

I think I can speak for AHRQ, even though I'm a contractor, in saying that the agency is very supportive of the STS registry and the development of outcomes measurement

programs based on STS and other registries, but it is the availability of alternative measures based on administrative data that sometimes encourages people to participate in registries.

So within certain local markets, users have the option of referring to either their local administrative data set or asking hospitals to join the STS, and so if it's understood in a local market that purchaser coalitions, for example, will produce a report card based on the AHRQ measure, then what often happens is that hospitals voluntarily say, "Well, we'd rather release our data through the STS," and that's perfect.

So if that alternative mechanism exists, then, if anything, it tends to increase participation in the registry, and in the long run the goal is to get information into the public domain that's useful to stakeholders. So I think we're all on the same page in terms of the pursuit of that

goal, and having complementary measures is actually a way of furthering that goal.

MEMBER WILHOIT: And I think it's really important to have the complementary measures, because I just looked up STS, and for Illinois we have 13 hospitals publicly reporting the CABG data. Well, 13 isn't very many of our hospitals that do the surgery.

You know, if pediatric isn't even in the loop yet on public reporting, it's obviously going to be a while before that information is available to anybody except the hospitals, so I think a measure that is available on a broader scale and that can be run on an administrative database, you know, so that everybody has the information is useful.

DR. BURSTIN: And one other question might be going forward it would be really nice to be able to see the complementarity or the similarities and differences between the data sources.

So I think it would also be very useful, for example, if STS began collecting as part of the registry some of that key claims data so we can begin to make sense of those connections, which I know has not been traditionally part of the STS registry. Again, anything we can do to learn more about how these different resources can get us complementary and/or contradictory information, I think that'd be really useful.

MEMBER ZAMBRICKI: I had a question for the surgeon from Boston University Children's having to do with the RACS statistical risk model. Did I hear you say that that is used worldwide and it's the largest risk model used in pediatrics or pediatric surgery?

DR. JENKINS: Well, that's certainly been my experience and perspective. By the way, I'm not a surgeon. I'm a cardiologist. I'm also the Chief Safety and Quality Officer for the Children's Hospital in

Boston.

So I'm sure that Jeff and others might disagree. I'm not suggesting that the methodology used by STS and their sister organization in Europe, the EACTS, is not also widely used, but they all have required participation in registries.

In the United States there was a survey that was done by a group of pediatric cardiologists about what was the preferable methodology, and three-quarters of the survey respondents actually chose the RACS methodology in the United States. The methodology has been used in South America.

It's being used in a large -- in a very large developing world collaborative which we're doing now for bench marking in developing world countries, and it's been used by the Children's Hospital Corporation of America to generate reports internally across all the large children's hospitals. So I do think that the comment that it's been widely

used is real.

CO-CHAIR TORCHIANA: So, I don't want to cut off any productive discussion, but it seems like we've arrived at a point where there seems to be pretty strong consensus that these are potentially complementary measures and that the -- asked that we would have ultimately, I think, was just described that it would be great to see, A, the STS measure in the public domain as that evolves and, B, that the two measures be used in a comparative way on registry and clinical data or administrative data to try to work out any kinks or irregularities. Otherwise, I think we're on pretty sound ground here, unless somebody wants to raise another issue.

MR. JACOBS: Can I just make one very brief comment about -- first of all, I agree with what you said. I think that's a fantastic plan moving forward, and I just wanted to have the opportunity to put into the record a few facts about the size and scope of

the measure that we're proposing and its utilization.

So the largest congenital heart surgery database in North America is the STS database. The largest congenital heart surgery database in Europe is the EACTS database, and both of those two large-scale databases have unanimously endorsed the method that we're proposing.

Not only that, but when all of the surgeons on the STS Database Task Force were asked which methodology would they prefer to move forward with, all of those surgeons on the STS Database Task Force unanimously chose to endorse the measure that we proposed as a society.

Now, that speaks to the size, scope, and support behind our measure, but I do believe that these two approaches are very complementary and are additive, rather than competitive.

MEMBER DILLON: The question is

still going to be, though, that we still have to narrow this down from three to two. Is that still not the ultimate goal?

DR. BURSTIN: I think part of what we heard is that AHRQ and Children's Hospital Boston will have some ongoing discussions, and I think we'll provide you the additional information as that goes forward, and then we'll see whether -- you know, we'll line it all up for you and see if it all makes sense on an upcoming conference call.

MS. MURPHY: So, may I just ask a question of the group? Anticipating further discussion on a future conference call based on the fact that, one, you just got some background about how to assess related, competing, and you could say complementary, as well, and you have first heard the discussion of two of these measures, and, as has been pointed out, you have not seen the documentation for two of the measures that you've seen of one, are there materials that

you would like to see to help better be informed to come to a point of recommendation of suggestion about how to go forward, and do you have any other questions of the developers that you'd want to have answers to in preparation for a future discussion?

CO-CHAIR TORCHIANA: I guess I'd answer that by saying the ideal thing to see would be just a cross-comparison of the two data sets on the same population with whatever level of detail is available.

That's a pretty big project, but that obviously would be the most definitive contemporary way to look at the performance of the two data sets and determine to what degree the complement and to what degree they contradict.

DR. JENKINS: This is Kathy Jenkins. One of the -- I think the only group that could really do that is probably the STS, because the quoting framework that they used isn't really used by anyone else.

MEMBER DUTTON: My request would be simpler. I'd just like to see from the STS the, I guess, the bin size. So, by facility, by category how many cases do they have for a year or two years or three years?

MR. JACOBS: That information has actually been published by STS in the Journal of Thoracic and Cardiovascular Surgery, so there's a manuscript that's referenced in the material that was distributed to the group that publishes that exact data.

MEMBER DILLON: David, what you're asking for, I think what we would want is, you know, for whatever, a complementary or a year report is for, you know, one, two, and three, which would allow us to compare institution to institution across the different reports. Is that correct?

CO-CHAIR TORCHIANA: I think when it comes to competing measures that would obviously be as close as one can get to a gold standard. Now, to what degree that actually

identifies the best in breed, I'm not sure, but it is -- it has been done with the AHRQ measures in adult heart surgery against the STS measures, and it's useful information. It gives you an idea of where they coincide and where they diverge.

MEMBER DILLON: I would -- I would agree. I think even if they're -- there will be separate reports, but it would allow us as Committee members to look at the type of or how the institutions are reported out and how to correlate those results.

CO-CHAIR TORCHIANA: I guess I'd say that's a big project, depending on what time line we're on, and I feel fairly satisfied that these measures should be approved as complementary, but that work really would be very helpful.

MEMBER HALPERN: Do we think that the AHRQ and the Children's Hospital should be essentially one?

DR. JENKINS: I just want to

remind everyone that the Children's Hospital measure was proposed for administrative and non-administrative data, as it was originally validated.

MR. ROMANO: There are some procedures that National Quality Forum has, I think, with regard to the identification of measure stewards and how we would establish a co-stewardship that I think would have to be worked out.

I can't speak for the agency on that particular issue, because there would have to be a specific agreement by which we would have joint responsibility for maintaining the measure and for keeping NQF up-to-date with respect to new evidence about the measure and changes in the indicator specification.

One of the disadvantages of administrative database measures is that the definitions have to be reexamined every year as new ICD-9 CM codes are introduced or

removed. We're also looking forward to conversion to ICD-10 CM in October 2013, so all of these measures that use codes will have to be respecified, and there will be a lot of ongoing dialogue with NQF about that process.

So, in any case, we're happy to work with NQF and with Dr. Jenkins and her colleagues to bring these into a single measure, recognizing that there may be some temporal issues associated with the sort of legal issues and working out the software compatibility.

MEMBER KLEINPELL: I guess I have more of a general question. It's not related to these three measures specifically, but as we move forward, when a measure developer is submitting a new measure that may be a competing measure, are they required to submit a rationale for why they're submitting it? Otherwise, we're just going to see a proliferation, I think, of measures being submitted and having to make decisions.

DR. BURSTIN: Yes, actually, our updated measure submission form requires as a condition of submission that you've looked in the NQF portfolio, you have identified what else there is, and, in fact, we're going to be -- we're actually going to have the advantage of being able to announce projects significantly in advance of the due dates.

We will actually expect by the time it's submitted to us that the developers would have done the work of looking and harmonizing this. As you guys know, it just takes a long time, and doing it in the course of a project just delays things significantly.

We're also going to try to, as much as possible, let everybody else know we're working on some pipeline things, as well, so people can go, we hope, ultimately somewhere to be able to say, "These are the developers who are working in my space," and try to, as much as possible, work collaboratively.

MEMBER HALPERN: Is there any move to make a global database that everybody can access various data pieces that fulfill everybody's need? Do you understand what I'm saying?

DR. BURSTIN: I'm not sure I know what data pieces are.

MEMBER HALPERN: Like a huge registry, basically, a huge registry of data that --

DR. BURSTIN: Of measures or of data?

MEMBER HALPERN: That everybody can, you know, use based on what everybody's needs are so that, you know, everybody is -- because the thing that keeps coming up is that everybody's data points are slightly different.

DR. BURSTIN: There is some of that work going on as part of the work NQF and others have been doing on something called the Quality Data Model, where they have been

trying to identify the key data elements that would populate electronic health records in particular, and we hope registries, to be able to ultimately do quality measures that we know we need and want.

Part of that will ultimately be, and there is still some work being done to figure out where this will reside within HHS, of who would be the code set owners, which I think is part of what you're trying to get at.

So if they're identifying these procedures this way, can somebody kind of take that same list the next time they're developing measures so it's not this constant churning of figuring that out? That's really just starting.

There is some interesting work -- actually, Patrick probably knows some of this -- funded through AHRQ, actually, by a group called USHIK, U-S-H-I-K, which actually does pull together a lot of the existing data and data sets to help do that.

MEMBER HALPERN: So are they involving the societies of the various -- because what keeps coming up from the different societies or the different specialties is that we on the clinical side view things a little bit differently than those on the administrative side, so, you know, to blend those two sides together in terms of a -- you know, again, the idea is to, one, make quality better and also to be able to give to the patient something that they can view and understand.

DR. BURSTIN: Certainly on the EHR side a lot of that is happening. There is significant outreach to try to figure out what are those data elements that would be incorporated into EHRs going forward, but it's just beginning.

MEMBER MORTON: There is a movement afoot about trying to harmonize all the different surgical specialties, Surgery Quality Alliance, and have been meeting for I

think almost five years now in the hopes of creating, you know, a unified surgical database. Efforts have not been fruitful to date in terms of getting a single data set out.

MEMBER DUTTON: AHRQ is working on a registry of patient registries right now that would look like -- seriously, ClinicalTrials.gov.

MEMBER MORTON: Department of Redundancy Department?

MEMBER DUTTON: That would list all both quality management and research databases that existed. It's at the white paper stage right now, so they really haven't gotten very far, but the idea would be you could drill into it, see what elements each registry is collecting, and see what data is out there.

DR. BURSTIN: And there are some great international examples like Sweden, where they actually have sort of done the

registry of registries that pulls it together. That would be lovely.

MEMBER HALPERN: Sweden has that massive database, and anybody can use, and they have like hundreds of data elements, so each society can pull from it and get whatever information that they want.

CO-CHAIR TORCHIANA: Sure. All you need is the Swedish health system with a single government payer, and we're there. I think we've completed that agenda topic.

I'd like to thank our developers, and I think we can move on to the next topic on the agenda. So that would be NQF member/public comment if anyone is on the phone. Is there anyone on the phone and muted, as there was yesterday? We're a little early, I guess. Okay, we'll go to Melinda for next steps and time line.

MS. MURPHY: So, Alexis and I talked a bit earlier, and what we believe is that based on the fact that there is some

information that you wanted to have back regarding some of the measures that were discussed, that we need to get that collated, get it out to the developers, let them provide the information back, much as was done with Phase I, and give you the opportunity to review and react to whether or not the information that's provided meets your specifications, your conditions, and do that as step one in a conference call and then soon thereafter follow up with a second conference call to have the discussion about related and competing measures, related and competing measures from Phase II, and we can close out the discussion on the one that you've just begun.

So we will have to get you cleaned-up versions of that table with related and competing that you looked at a bit earlier, so that will be the second of two conference calls the way that we've talked about in terms of volume of material to be

covered to do, and we'll get out to you a request about time availability to do that in two calls. So that's next steps.

CO-CHAIR MORRIS: All right. So just, in closing, I'd like to say thanks so much for all of your hard work and attention. I think everybody really contributed substantially to the meeting, and, of course, we're not really done yet, so there's more coming. I'd like basically to applaud the Committee for all of the hard work and also the NQF staff.

DR. BURSTIN: And I would like to thank everybody for your patience with NQF processes and evolution here. We are trying to make it a stronger process, and we have been doing a little bit of flying it while -- what's that expression? Building it while flying it, so thank you for your input, and thanks to our Chairs for the great effort.

(Whereupon, the foregoing matter was adjourned at 2:31 p.m.)

J

N

S

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Surgery Endorsement
Steering Committee

Before: NQF

Date: 05-05-11

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter