NATIONAL QUALITY FORUM

Moderator: Christy Skipper
November 7, 2016
2:00 p.m. ET

Operator: This is Conference #46828005.

Operator: Welcome, everyone. The Webcast is about to begin. Please note today's call is being recorded. Please standby.

Christy Skipper: Good afternoon, everyone. And welcome to the Surgery Post-Comment Call. My name is Christy Skipper, project manager for the Surgery Project. And I want to just turn it over to folks around the table at NQF to introduce themselves.

Katie Streeter: Hi, everyone. This is Katie Streeter, Senior Project Manager.

Marcia Wilson: Marcia Wilson, Senior Vice President, Quality Measurement.

Elisa Munthali: Elisa Munthali, Vice President, Quality Measurement.

Helen Burstin: Helen Burstin, Chief Scientific Officer.

Christy Skipper: Thank you. Before we get started, I'm going to go ahead and take roll, and then we'll hear from any opening comments or remarks from our co-chairs.

So, when you hear your name, please say, "Here". OK, Lee Fleisher.

Lee Fleisher: Yes, here.

Christy Skipper: William Gunnar.

William Gunnar: I am here, thank you.

Christy Skipper: Karl Bilimoria.

Karl Bilimoria: Here.

Christy Skipper: Robert Cima. Richard Dutton.

Richard Dutton: Here.

Christy Skipper: Elisabeth Erekson.

Elisabeth Erekson: Here.

Christy Skipper: Frederick Grover.

Frederick Grover: Here.

Christy Skipper: John Handy. Mark Jarrett. Clifford Ko. Barbara Levy.

Barbara Levy: Here.

Christy Skipper: Barry Markman.

Barry Markman: Yes, good morning.

Christy Skipper: Hello. Kelsey McCarty. Lawrence Moss. Amy Moyer.

Amy Moyer: Here.

Christy Skipper: Keith Olsen. Collette Pitzen.

Operator. OK. Sorry about that. Collette Pitzen.

Collette Pitzen: I'm here.

Christy Skipper: OK. Lynn Reede.

Lynn Reede: I'm here.

Christy Skipper: OK. Christopher Saigal. Salvatore Scali.

Salvatore Scali: I'm here.

Christy Skipper: OK. Allan Siperstein.

Allan Siperstein: Here.

Christy Skipper: Larissa Temple. Melissa Thomason. Barbee Whitaker. And A.J. Yates.

A.J. Yates: Here.

Christy Skipper: OK. All right. Prior to this call, we've sent out the commenting memo as well as the comment table. I want to ensure that everyone has those materials in front of them today.

And if you have not, please call out your name and we'll resend it to you.

And also, I just want to be sure that you all logged in to the Webinar today through the link sent by (Sean Vettori). So, can everyone see the slide?

And if you have not, call out -- if you can't, call out your name.

OK, so it sounds like...

Male: I did that, we're on the -- we're on the page that I see the flag.

Christy Skipper: A flag?

Male: No, I -- you said...

Christy Skipper: I'm sorry.

Male: I misunderstood you, I thought you said we had to see something.

Christy Skipper: You should see the first page of the comment memo on your screen.

Male: OK, I thought there was something in reference to voting.

Christy Skipper: That'll be later on, yes.

Male: OK, good enough.

Christy Skipper: OK. So, it sounds like everyone has their materials. So, I just want to let you all know that the purpose of this call is to review and discuss comments received during our public and member commenting period, and then also to revote on two measures where consensus was not reached.

We also received a request for reconsideration, so the committee will be reviewing that request and deciding how they would like to move forward. And then also, we will be looking at the comments received and determining a response from the committee.

And without any other comment, I'll turn it to Lee and Bill for any opening comments or remarks and then we'll get started with our agenda for today.

Lee Fleisher: Sure, if you'd like me to go first. It's Lee, and I wanted to thank everybody for their continued strong efforts and particularly thank staff for putting together an excellent memo and the responses.

In reading through it, I just wanted to remind the members of the committee that we're asked to review different measures based upon the NQF criteria. And I think some of the -- one of the advantages having both Marcia and Helen on the call today, if there's questions about that, because that's some of the comments that we saw from -- after our report is about that, we can be sure that we appropriately look at whether or not we judged the criteria correctly or whether we evaluated some of these measures outside of that criteria.

And I think that'll become clear as we go through it. But I wanted to thank you for -- because I think the memo nicely clear our thought that staff created, as we clarified some of those issues.

Christy Skipper: Thank you.

William Gunnar: Yes, this is Bill. I have nothing else to add, so carry on.

Christy Skipper: OK ...

William Gunnar: Do we have -- by the way, do we have quorum?

Christy Skipper: Yes, we do, we now have quorum.

(Off mic)

Christy Skipper: So the first order of business is the review of the reconsideration request for measure 0351, death among surgical inpatients with serious, treatable, complications.

Lee, are you there?

Lee Fleisher: Yes.

Christy Skipper: OK. All right, so we're moving onto our first agenda item, the reconsideration request for measure 0351.

Lee Fleisher: OK, do you need me to -- I'm sorry.

William Gunnar: So the other -- this is Bill, let me jump in. Do we have the two committee members that originally reviewed this and presented it at the meeting?

Christy Skipper: Right, so, two of our discussants, Christopher Saigal and Amy Moyer are on the phone. And actually, I want to call out to, I believe, Patrick Roland ...

Female: Romano.

Christy Skipper: Romano...

Lee Fleisher: Romano.

Christy Skipper: ...if he would like to give us an overview of their reconsideration request.

Lee Fleisher: Great. Patrick?

Is he on an open line?

Pam Owens: So this is Pam Owens...

Patrick Romano: Oh, I am, yes.

Pam Owens: ...from AHRQ. Oh, good, Patrick, you're there.

Patrick Romano: I'm UC Davis. But let me defer first to Pam Owens from AHRQ.

Pam Owens: So, I just want to thank the committee for entertaining our reconsideration request. We did put a memo together and we feel that this is a long-standing maintenance. We put it in for maintenance endorsement, this is a long-standing indicator that has been endorsed by NQF since 2008. Because it is a maintenance indicator, we expected the review to follow a maintenance submission criteria as outlined by NQF as -- in 2015 and 2016 document.

So our reconsideration request specifically falls around our discussion around validity. Patrick can speak more about that in just a moment. It also into used case of the measure prior to a full discussion of scientific acceptability and also making sure that as we discuss the used case that the used case is agnostic to its particular use but rather is it usable for quality improvement and to -- quality improvement and accountability purposes.

And then, the third is that just to note that there was a related but not competing measure that did pass in the Patient Safety Committee just a few months prior just so that there is consistent review across committee.

So Patrick, I'll let you talk about PSI 4 and specific more details around our -- or asked regarding reconsideration for a revote.

Thank you.

Patrick Romano: Yes, I'll be very brief. Just to call people's attention to the detailed memo that is at the end of the document that you received from NQF staff, it's the document labeled October 21st with additional information regarding maintenance measure NQF 0351.

Basically, this document just summarizes our response to the key issues that were raised in the last Surgery Standing Committee meeting. And, I would just point out that we did -- the discussion and the concerns that were raised did prompt us to go back and reassess the risk adjustment model. And, we do think that we could incorporate enhancements to the risk adjustment models for PSI 4 to better account for the concern the committee members raised regarding the fact that some patients are transferred from one hospital to another with complications that occurred outside the second hospital.

So, in order to account for that concern, we've re-estimated risk adjustment models, including variables in each of the five stratified models to account to -- or whether the triggering complication was present on admission, and whether that complication was relatively severe or relatively mild.

And, there are tables attached to the appendix that demonstrate the impact of that. So I think that with those changes, we do see some significant improvements in the performance of the risk adjustment models. So they're showing c-statistics ranging from 0.715 up its highest, 0.878, which I think are excellent c-statistics for these kinds of risk adjustment mortality indicators.

We also addressed some of the specific concerns related to the testing data. Basically, there's a very robust amount of testing data both from prior published research as well as from our own work, which is summarized in this document as well as in our original submission. So I think properly considering the fact that the validity standard for measures of this type is based on the validity of the computed performance measure score and demonstrating that the computed performance measure score at the hospital level in this case is correlated with other structure and process measures really meets the NQF criteria for acceptability on the validity standard.

So, we're happy -- I'll turn it back to committee members and we'll be here for any questions that may arise.

Lee Fleisher: Great. So, I guess it's Amy, who were the two?

Christy Skipper: Amy and Christopher Saigal.

Lee Fleisher: Great.

(Multiple Speakers)

Amy Moyer: So this is Amy. I'm actually looking through our original notes in the original documentation of the validity algorithm that have been in the measure worksheet. I would have passed this measure on validity originally on consistence with that algorithm. It is good to see the changes that have been made that have even increased the validity beyond that and I believe that we should reconsider it.

Lee Fleisher: Great. Other comments or questions, or?

Christy Skipper: Are there any comments from the other lead discussant, Dr. Saigal?

OK.

Female: I'm not (technical difficulty). I actually got an e-mail from him this morning that said he was delayed in the operating room -- on the call.

Christy Skipper: OK.

Female: OK, thanks.

Christy Skipper: Thank you.

Lee Fleisher: Yes.

Christy Skipper: So, hearing that you would -- all would like to reconsider, we are...

(Multiple Speakers)

Female: Yes.

Lee Fleisher: I really want to thank Patrick for going back and re-looking at the variables and re-looking at the model based upon our concerns.

I also agree with Amy that the measure developer was extremely responsive to some of those concerns.

Frederick Grover: Yes, I agree with that comment as well, Lee. This is Fred.

Female: So do you need a motion to reconsider? If so, so moved.

Christy Skipper: Yes, OK.

Female: Any second?

Lee Fleisher: Seconded.

Christy Skipper: OK. So, it sounds like the committee wants to reconsider.

Operator, can you provide instruction for the committee to revote on -- or to vote on validity?

Lee Fleisher: So we can vote in real time?

Christy Skipper: Yes, so, you should -- now, on your screen, you should see a slide that says, "Surgery Post-Comment Call", and I'm going to advance to the voting slide for validity.

Operator: And as you advances to the voting slide for validity, you will notice some boxes to the side of the individual choices next to the A, B, C and D. Just click in the box next to the answer of your choice and it will register your vote and the reporting will show your name with the corresponding vote.

Male: Oh, is that the only discussion we're having on this?

Christy Skipper: So...

Lee Fleisher: Would you like -- sorry.

Christy Skipper: So if you all pass the measure on this criteria, we would move on to vote on feasibility and usability and use, and an overall recommendation for endorsement.

Female: And we're still showing 16 as our number, so we're just missing a few votes.

Female: So did others on the committee have any dissenting views or didn't feel like we should reconsider, we'd love to hear your input now.

Male: I mean, there was a lot of discussion at the meeting about validity issues well beyond what we just discussed, and whether it was just transfer status and whatnot. So I think -- I mean, a number of people made a number of points that were related to the underlying data source, the validity of the actual components of this measure, and -- I mean, just fixing a couple of these things, I don't think, changes a lot of that.

So I'm not sure that, you know, diving into a reconsideration without sort of opening up all of this again is really merited.

Richard Dutton: This is Rick. I'll note that the transfer issue which concerned me was appropriately addressed by the developer in particular doing the sensitivity analysis that showed that including or excluding transfers would have very little effect on the outcome of the measure.

Barbara Levy: This is Barbara, and I agree. I think that was the major issue was the issue of transfers and having resolved that and done an additional calculation, I'm comfortable.

Lee Fleisher: I think there was a discussion, this is Lee, about the data source and while people -- and I -- am I correct in that this was the one -- this was the measure in which we had a discussion about administrative versus clinical data?

Female: Yes, you're correct.

Lee Fleisher: And -- but I think that the respondents and others can chime in appropriately with regard to the criterion. Helen, maybe you can -- or Marcia can also comment that this is the measure were brought forward and while it would be great in the future to have data from an HER, that the current time, this is what we have and that's what (we'll be asked) against the NQF criteria.

Was that the big concern that was just raised?

Allan Siperstein: Hi, Allan here. Yes, no, I agree. I remember and I've got notes on specifically that aspect of things at the collection of administrative data make

this a more useful measure to longitudinally track your own hospital. But, raises questions in terms of being able to compare hospital A to hospital B based on, you know, how well they trap that administrative data.

Lee Fleisher: Helen or Marcia, could you comment about data -- the criteria and how we should look at that?

Helen Burstin: This is Helen. I mean, there's nothing on face value that says one data source versus another, it has to be factored in overall into how it reflects the testing on reliability and validity. And I think the question before the committee really is, have the responses of the adjustments made by AHRQ potentially change the way you would vote on the measure as is. Obviously, you have to vote on the measure in front of you not -- you know, not specifically that's on the question of data source, but what are the results of testing suggest about the reliability and validity of the data source.

Collette Pitzen: This is Collette. I have a question and a clarification. It seemed to recall a lot of discussion as people have indicated about the issue of patients who are transferred to another facility and then that second facility kind of bearing both (from) (technical difficulty.

And if I'm understanding correctly, now the developer is considering to risk adjust by the instances of transfer which, in my mind, says that that takes into account some of the tertiary aid care centers that would be receiving a lot of transfers.

So, that in my mind helps with that whole transfer issue, but I guess I'm just asking for clarification if that's now part of the risk adjustment model.

Female: Patrick.

Patrick Romano: Yes, that's correct. The risk adjustment models for PSI 4 have always adjusted for whether the patient is transferred in, that is the source of the patient. But now, with the changes that AHRQ is planning to implement for version 6.02, we'll also adjust for whether the triggering complication was

present on admission, in other words, whether the patient arrived at the hospital with that complication already present.

And we'll also adjust for the severity of that complication, for example, whether it was a deep vein thrombosis for a -- versus a pulmonary embolism.

Collette Pitzen: Great, thank you. For myself, that clarifies some of the issues I had around validity related to the classifications themselves and for risk adjustment.

Male: So I think one of the other issues we talked about was the issue of DVT and P.E. being included in here. So, if there is a surveillance bias associated with detection of that event, hospitals that detect more will have more cases in the denominator than others.

And, I mean, there is a -- there's an argument to be made that you could exclude DVT, P.E. and still have a useful measure, right?

Christy Skipper: Would the developer like to respond to that?

Patrick Romano: Sure. I mean, again, the concept behind this measure is that high-quality hospitals that are performing well have effective multi-disciplinary teams that include physicians, nurses, therapists, others that can identify these complications early and intervene effectively to prevent the patient from getting into worse shape and ultimately dying.

So, that concept has been validated many times in many different ways. And I think that concept applies just as well to DVT P.E. as it does to other measures. In fact, if you play out this concern about surveillance bias that you're describing, if a hospital, for example, diagnosis more early stage DVTs, then in fact its PSI 4 rate will drop, right, because it will find thrombosis that -- and may be able to initiate treatment early and in fact, that could be a way to lower its PSI 4 rate.

So, in general, the things that lower PSI 4 rates will be interventions that focus on early identification and prompt and aggressive treatment of complications that arise in the course of hospital care.

Barbara Levy: And how are we, Patrick, avoiding the gaming in which a major complication occurs, how are we tracking the hospital that transfers high patients (out). In other words, I am comfortable now with the receiving hospital (data) adjusted and not being judged inappropriately. But what are we doing, how does this measure account for those hospitals that are gaming the system by transferring those patients out before they die?

Patrick Romano: Well, there are two things I would say about that. So, one is that the great majority of the hospitals that do that frankly are small hospitals, often rural hospitals. Many of us who work in teaching hospitals receive those transfers. We know about that process.

And, those hospitals generally have such low volumes that their risk adjustment outcomes are indistinguishable from the national average. And that's just the nature of the process. So, when you look at those small hospitals, their smooth rates end up, you know, being essentially equal to the national average. So that's an inherent problem with all quality measures as they're applied to very small hospitals, rural hospitals, so forth. So there's nothing unique or different about this measure in terms of how it deals with that, it's an inherent problem in the outcome measurement, the enterprise.

The other point I would make is that we do have -- most of the conditions that drive the denominator of PSI 4 also show up as numerator events for other patient safety indicators. And that's by design.

So, that if a hospital has a high rate of post-operative complications, then it will tend to show up with higher rates of those PSIs, such as PSI 12. So, there's a little bit of a countervailing force, if you will, in that people are looking at the same data from a variety of different angles including looking at the incidents of those operative complications. As well, there's of course

looking at the hospital payment that's driven by those post-operative complications.

Frederick Grover: This is Fred Grover. If I can ask you one more question, how granular or to what degree of specificity do you have for the admission to the second hospital in terms of the acuity of that patient? I mean, there are hospitals that could be in good faith forwarding patient, referring patients fairly quickly when they see they have a potentially fatal complication. They still come in relatively stable. Whereas, there are others, a way to absolute last minute to refer somebody who's basically moribund when they get there. What are your levels of acuity when they come in to that second hospital?

Patrick Romano: Well, it's a valid point that these types of measures based on administrative data including all of our, you know, risk-adjusted mortality measures, risk adjustment complication measures, readmission measures, so forth, are all potentially subjective to what you might call confounding bias or omitted-variable bias, because the administrative data don't have detailed information about physiologic severity of illness.

Obviously, we can adjust our factors like the patient's age, the patient's -- whether the patient was transferred in, what complications they had. We can also adjust for all of their comorbid conditions like heart failure and COPD, and asthma and so forth, renal disease. We can adjust for the type of operation that they had, and the magnitude or complexity of that operation.

But, in terms of, I think, what you're describing which might be about, you know, what the patient's blood pressure was when they hit the door. That would be difficult to capture frankly with administrative data.

So we have to rely on other proxy measures and we have to rely on the evidence regarding what we call construct validity and that we know from the studies that are cited here and elsewhere that in general, hospitals that have excellent nursing teams that have high-nursing skill mix, that have more qualified medical staff, these hospitals on average do better on or in similar measures.

Frederick Grover: Yes, I -- that's where I was getting at, it would be something to say if they come in in shock or they come in on a ventilator, would that not be captured administratively?

Patrick Romano: Good question. I think that...

Frederick Grover: Or is it come in with septicemia? I mean, something is fairly catastrophic. And I'm not talking about a specific blood pressure, but what one would call a shock. You know, I realized that even that definition is somewhat subjective, but.

Patrick Romano: Yes. I think -- no, I think that that's exactly right and we try to get at that with what we described as the severity of the complication that the patient came in with. So, for example, if the patient came in with sepsis, then we have -- now, we have a flag in the model to indicate if they were in septic shock, or if they were in sepsis without shock.

If they came in with thromboembolism, we have a flag indicating whether they came in with a P.E. or DVT. If they came in with GI hemorrhage, we have a flag indicating whether there was evidence of a perforation or not with that hemorrhage.

So, in each case, we've tried to create a severity flag to capture that as best we can with the administrative data exactly as you're describing, Dr. Grover.

Frederick Grover: OK. Thank you.

Christy Skipper: OK. I just want to remind you all that we -- if you all decided to vote on validity that we do have two more criteria that you will need to vote on following that. So, if there are no other comments, just to hear again that the committee want to accept the reconsideration request, and if so, then we will need to move the vote.

Lee Fleisher: I'll move to the revote.

Barbara Levy: I agree.

Christy Skipper: OK. OK. So, it looks like in front of you, you can see that we've had 14 people vote so there are at least two other individuals who have not voted, so if you could just take this time to submit your vote. We're voting on validity for measure 0351.

We have 15 votes, which is quorum. OK.

Now, we have 16 votes. Voting is closed. 25 percent vote -- 27 percent vote high, 60 percent moderate, 13 percent low, 4 percent insufficient. The measure passes on validity.

Now, the committee will discuss feasibility. And again, the lead discussant, Amy Moyer and Christopher, if there are any comments you have and then we can discuss and move to vote.

Amy Moyer: This is Amy Moyer. The only comment I had regarding feasibility. This is a measure our organization calculates and generates and we know several other organizations that do as well. And, this is very straightforward to use and supplies our Q.I.s operator to generate measure results. And the data sets are very available.

So, our experience with this (is that it's) very easy to calculate and use.

Christy Skipper: Any other thoughts or comments on feasibility?

OK. We're now voting on feasibility for measure 0351. One high, two moderate, three low, four insufficient.

OK. All votes are in. 38 percent vote high, 63 percent vote moderate, 0 percent low, 0 percent insufficient. The measure passes on feasibility.

Now, onto usability and use.

Amy Moyer: This is Amy. This measure is widely used in variety of programs, so I would concur with either moderate or high rating that was on the measure worksheet.

Christy Skipper: We're now voting on...

Male: Well, they have been used because it's had NQF endorsement. But I think that there are, you know, sufficient number of people questioning whether it should. And, I don't know that that should be the case. And I think the discussion that we had was that if you consider the use of it as in public reporting, most of us decided probably wouldn't be useful for comparing hospital quality.

So, I mean, the fact -- if we're supposed to be agnostic to the use but yet here, still talking about public reporting within six years, it does seem like we are being instructed to talk about use. So, I'm a little confused by that.

Frederick Grover: Maybe you could -- I -- this is Fred again. I have a question on the improvement. Maybe you could just comment on that for the ...

Amy Moyer: So this is Amy. Really quickly on the usability and use, I guess my take on this is that, if a measure were to be endorsed, but yet was not in use in any kind of a meaningful program, it could lead us the question whether that measure really is feasible or really is of importance to individuals who are doing the measurement. The fact that it is out there and being used kind of, I guess, confirm that in some ways, the importance of the measure or the fact that it is actually feasible and able to be used.

Allan Siperstein: I'll agree with some of the prior comments here. And in talking to some of my own quality people around here who actually collected at and looked at the data and analyzed the cases, you know, this is one of the measures that they're -- kind of the least comfortable with in terms of passing the, you know, is it clinically relevant sniff test. And when they look at individual cases, often kind of see a disconnect between kind of a trivial initial procedure that's done

and an ensuing what, you know, what they think is of unrelated kind of train of complications that have happened.

And it's not just the one exception to the rule but just kind of a feeling that given the matrix of complications that are trapped and procedures that are trapped that there are a lot of these cases that are collected that may not be the world's most clinically relevant.

(Multiple Speakers)

Allan Siperstein: But...

Male: Yes.

Mark Jarrett: This is Mark. Because I lived in the quality world and I tend to agree, I think of, you know, a lot of the PSIs as problems but this one in particular is very often not specific enough that really gives you information which lets you really, you know, help in terms of performance improvement. It's not going to generate answers that -- or data that's really going to direct you because it's kind of diffuse stuff that you're collecting and it's very hard to pick up patterns.

Christy Skipper: Are there any other comments on usability and use?

Male: I think those comments are right. I think most people -- I think most hospitals are worried about this measure and say that it has probably the least face validity when they go and dive into their cases of any these.

So, I think it's worrisome to be able to pass this measure. I think it really hurts the validity of the NQF process if we keep passing these measures, yet most of the world seems to be complaining about the measure even if it is used, it's used because it's one of the few things left that's -- that are available nationally, but that doesn't mean it's right. And then I think there are plenty of people who are advocating that we abandon PSIs altogether.

Patrick Romano: This is Patrick. Could I respond?

Lee Fleisher: Please.

Patrick Romano: Yes, I think it's important for people to keep in mind that this is the risk-adjusted mortality measure, and NQF has a portfolio that includes many risk-adjusted mortality measures for a variety of different conditions and procedures including M.I., pneumonia, heart failure, stroke, et cetera.

And, we never argue that all of these deaths are preventable, certainly, hospitals, each of our hospitals has their own internal process for reviewing deaths and assessing the potential preventability of the death. But a measure like this, any kind of risk-adjusted mortality measure is most valuable when it's used as a rate, and when the rate is tracked overtime, when the rate is compared across units.

So, we actually discourage people from focusing too much on individual cases and encourage people more to look at rates and how those rates track overtime across institutions, across units, and what kinds of interventions and systems can be put into place to reduce the rates because it's very clear that there are things that hospitals can do to lower their rates. So that's really where the focus of attention that this indicator should be rather than on nitpicking each individual case.

Allan Siperstein: I like to argue at the same time, you know, as measures have been, you know, road tested for a number of years, they either kind of passed or failed the sniff test in terms of, are they really clinically relevant in identifying correctable problems.

Lee Fleisher: Are there other comments? We've heard, I think, differing opinions.

No. Then, should we still vote and see how the group...

Female: Yes.

Lee Fleisher: ...thought comes out on this?

Christy Skipper: Yes.

Female: Yes.

Christy Skipper: So, we'll be voting...

Male: Yes.

Christy Skipper: ...on usability and use. One high, two moderate, three low, four insufficient.

Voting is closed. 13 percent vote high, 25 percent moderate, 63 percent low, 4 percent insufficient. The measure does not pass on usability and use. However, we can still move on and take an overall vote on suitability for endorsement. One yes, two no.

Lee Fleisher: So we're still missing one vote.

Christy Skipper: Yes, one more vote.

Male: Maybe they're voting for the green party.

Christy Skipper: So, maybe the person stepped away, but we have 67 percent vote yes, 33 percent no, measure 0351 is recommended -- pass its overall suitability for endorsement.

Lee Fleisher: And -- go ahead.

Elisa Munthali: Sorry, Lee. This is Elisa. One of the things we neglected to mention, if you are speaking, if you could please announce yourself so that we can credit your comments accordingly. Thank you.

Lee Fleisher: And the comments that were made will go forward, correct, to the...

Elisa Munthali: Yes, in the report. Yes.

Christy Skipper: OK, moving onto our next agenda item. So we will need to review and vote on measures where consensus was not reached. So, we left off with measure 1543, Postoperative Stroke or Death in Asymptomatic Patients undergoing Carotid Artery Stenting. In this measure, you all cannot reach consensus on the evidence or validity.

So, we'll start out with our discussant, Salvatore Scali and Richard Dutton, once the developer provides any opening comments regarding this measure.

So, if -- do we have a developer for the Society for Vascular Surgery on the call?

Karen Woo: Yes. This is -- hello, this is Karen Woo.

Christy Skipper: OK. We are discussing measure 1543. Are there any comments that you would like to provide in regard to evidence and validity on this measure?

Karen Woo: Yes. Well, we believe that this is an important measure because the -- because even though CMS has not -- or only endorses or provides payment for asymptomatic carotid stenting in certain cases, it's still being performed, and I think by a variety of different specialties and I think it's very important to track this as an outcome, especially since it's a very controversial procedure that's still in the process of being tested.

The CREST-2 trial is going to give us a lot more information about it. But especially since it's still a somewhat controversial procedure, the outcomes for the procedure needs to be excellent in order for it to be performed and for the patient to derive a benefit from it.

Christy Skipper: OK. Discussants, any comments or questions for the developer?

Salvatore Scali: I agree -- this is (Sal) Scali. I agree with Dr. Woo's assessment. One of the things that's sort of the focus of the discussion were the apprehension about

the evidence, there's a variance (in) passion sort of rebuttal that was offered by Dr. Rosenthal regarding what they thought the discussion surrounded. Most of the concerns were the fact that societal guidelines aren't uniform. There are several guidelines that actually at the present time that don't endorse asymptomatic carotid stenting unless they're under the offices of a trial or within a registry.

And, I agree that this activity is ongoing. And one of the conundrums was that we face as a committee was, was there a precedent where we've previously endorsed quality measure on a procedure that's not necessarily universally covered by Medicare, because Medicare recognizes the controversy that exist with the current procedure.

I do agree that this is ongoing. I do agree this -- that the assessment of the literature that Dr. Rosenthal provides, I agree with many of these points actually. And I think we went over all of those issues regarding what the literature shows about compared to the (undergoing carotid artery) stenting. But really the apprehension about what the literature was, was just that because it's so controversial and still a procedure that's technically under study and we're awaiting the results (oppressed) to what would the position of the committee be to endorse the, you know, this important outcome measure for this procedure, on the procedures not necessarily universally covered by Medicare. That was really, I think, the core of the concern.

I do believe that we can get past that or we have to have a precedent where this has occurred previously that I wouldn't move to say that we should revote and I would recommend endorsing the measure if we can reconcile that point.

Richard Dutton: This is Rick Dutton. I agree with Dr. Scali. I think we should endorse this measure. Honestly, because it is a controversial procedure. We're not taking a stand on whether the procedure is valid or not. We're taking a stand that the outcome should be measured and we're offering a national-level endorsed and well-defined tool for measuring the outcomes of the procedure.

I actually think that endorsing the measure has a significant role to play in the discussion about whether and when the procedure is indicated. So, I favor endorsement.

Salvatore Scali: Agree.

Barbara Levy: This is Barbara. And I agree wholeheartedly. I think this is in fact a measure that really matters a lot and will drive appropriateness in care.

Christy Skipper: OK. It sounds like you all are ready to vote on evidence for measure 1543. One yes, two no. Voting is open.

Male: I don't see anybody yet. Oh, thank you.

(Off mic)

Male: You're expected to think outside of the box.

Christy Skipper: OK, just waiting on two more votes.

Female: They walked away, I believe.

Christy Skipper: All right, voting is closed. 80 percent vote yes, 20 percent no, measure 1543 passes on evidence.

Now, we'll be discussing the validity for this measure. I'll turn it over to the developer first and then the discussants.

Karen Woo: I'm sorry, this is Karen.

Salvatore Scali: Karen -- yes, go ahead, Karen.

Karen Woo: So, in terms of validity, I don't -- I think I missed the previous conversation regarding validity. Can somebody let me know what the controversy was?

(Brad): If I remember right, it was at the -- this was led by (Johnson) who was there at the day of presentation. The concern was maybe on the nine-month follow up. Since that time, we provided data (on that one) follow up and -- so there shouldn't be that much of a question on validity currently at this point.

Karen Woo: That's true. That's right. We have the long-term follow-up data. Unfortunately, I'm not in front of my computer. (Brad), can you provide that data?

(Brad): Yes, I'm not in front of mine. But, yes, really the long-term follow up is nine months. It did show the stroke rate was anywhere from 2 percent to 5 percent. But in 2500 procedures, it was quite well validated in that.

Salvatore Scali: This is (Sal) Scali. But I think that the measure -- there were no major issues with validity when we had the discussion. This was an in-hospital stroke and death outcome for asymptomatic carotid stenting, I believe it's the 1543 measure.

There were some concerns about whether it should be in-hospital versus 30-day in trying to harmonize the outcome measure with other sort of outcome measures, that was one of the sort of things that I have written in my notes was whether or not it should be harmonized...

Male: Yes.

Salvatore Scali: ...for this. But, yes, I don't think there was much any real concerns about the...

(Brad): Yes.

Salvatore Scali: ...methods or results of the validity testing that were presented at the time of the August...

(Brad): Yes.

Salvatore Scali: ...(rush).

(Brad): Yes. I think it was more -- it was just (all) mentioned early about the apprehension about approving procedure which was controversial. I don't think there was much of an issue of validity that day.

Amy Moyer: This is Amy. And I don't remember if this came up before, but I got -- the concern or question I have is that this is an outcome measure that is not risk adjusted it appears in any way. It feels like that could be a challenge to the validity of the measure.

Salvatore Scali: That was -- we actually had a pretty fairly significant discussion about this in August that it was felt that it shouldn't be risk adjusted due to the known relatively benign (indiscernible) of asymptomatic high-grade internal carotid stenosis...

Male: Right.

Salvatore Scali: ...to minimize stroke risk less than 2 percent in a (post-statin era). We didn't want to sort of "wash away" that decision making. And so, (our) patient selection was a critical issue and there was a concern that with risk adjustment that you may sort of mask that decision making to make or...

Male: Right.

(Multiple Speakers)

Salvatore Scali: ...effective but not risk adjust for this measure given the relatively benign natural history of the disease in the (post-statin era).

(Brad): Yes. That's what it was, so I remember that, yes. We have that discussion.

Male: Yes.

Amy Moyer: Thank you. It's coming back to me now.

(Brad): OK, and I will -- I'll try to remember August.

Christy Skipper: OK. If there are no other comments, we will move to vote on validity for measure 1543. One is moderate, two low, three insufficient.

OK, with 15 votes, 80 -- or excuse me, 16 votes, 81 percent moderate, 19 percent low, 0 percent insufficient. The measure does pass on validity.

(Off mic)

Christy Skipper: OK. And...

Male: All right.

Christy Skipper: And you all did vote on the remaining criteria, so we will now take a vote for overall suitability for endorsement. One yes, two no.

OK, with 15 votes, 87 percent yes, 13 percent no, measure 1543 is recommended for endorsement.

Male: All right, thank you.

Christy Skipper: The next measure that did not reach consensus is measure 3020, Initial Transfusion Threshold. If there are developers from the Joint Commission on the line, if you would like to begin to discuss your measure. I just want to note that the measure did not reach consensus on scientific acceptability and the feasibility criteria.

So do we have developer -- a representative from the Joint Commission on the line?

Michelle Dardis: Hi, this is Michelle Dardis from the Joint Commission. Can you hear me?

Christy Skipper: Yes.

Male: Yes.

Michelle Dardis: OK. Thank you. We did submit our comments in the public comment period focusing on recommending that NQF re-examine the process for evaluating for eMeasures submitted for approval for trial use and clarifying for standing committees and staff, as well as the measure developers the appropriate scope for the review.

We felt that the criteria we were preparing for, for approval for trial use were clearly met by the measure. Primarily, feasibility and testing which were not the focus of the discussion during the Standing Committee's deliberations and that that's really some of our comments.

Christy Skipper: OK. And our lead discussants for this measure were Barbee Whitaker and Lynn Reede. So, if either of you are on this call and have any comments on scientific acceptability.

Barbee Whitaker and Lynn Reede, any questions or comments or from any of the other committee members, you have anything to add on scientific acceptability?

Barry Markman: Wasn't there -- this is Barry. Wasn't there some special consideration for eMeasures? We had that whole discussion before that, that there was some other variables -- yes.

Frederick Grover: And this is Fred. Yes, as I recall, one of the major issues was in regard to scientific acceptability was that there are other factors that may well influence to trigger for transfusion such as hemorrhagic shock, bleeding, current active bleeding and so forth, which aren't really reported as part of this measure.

So it could be misleading and doesn't -- a couple of things that stand out in my mind. And I want to -- for -- thoroughly in absence of all those things agrees with having -- well, there are other soft factors too that the age of the patient, how well they tolerate being anemic in terms of their ability to get out of bed and move around and mobilize all of those types of things, which are chronic

lung disease status, are they hypoxic to begin with because of chronic lung disease.

I mean, there are all these patient variables that affect the clinician's judgment on whether to give a transfusion or addition to the hemoglobin.

Katie Streeter: This is Katie with the NQF. And just to clarify, this measure was to be considered for our approval for trial use status. So we won't be voting on a recommendation for NQF endorsement. Specifically, under scientific acceptability, what we're looking at are the eMeasure trials, measure specifications. And so that's what we'll be voting on.

It was noted that there were several concerns as you're discussing with the specifications of this measure. So we'll continue to gather your comments and then move forward to voting on those eMeasure trial measure specifications.

Allan Siperstein: The other issue that came out...

Female: Go ahead.

Allan Siperstein: Allan here. The other issue that came up when we had our former discussion was the tearing of the reported by hemoglobin level. That is, there was, you know, reporting for those less than seven between, you know, seven and eight, eight to nine, et cetera, and just felt that that was a somewhat un-interpretable scoring system to try to identify, you know, where the issues were.

Frederick Grover: Yes, that was part of what I was -- I think.

Female: Go ahead, Fred.

Frederick Grover: No, I was just saying that was part of what I was saying to rely only on the hemoglobin levels. I mean, in the absence of these other things, quite frankly, I always have in terms of patients are ongoing cardiac surgery, for example.

They all tolerated, but there are all these other things that go into that and the absence of capturing those, my concern and I think some of our committee at the time, and you all can correct me if this is appropriate discussion, you know, based on NQF criteria was that it's -- hemoglobin is an important factor but there are many others that are very important or it can even be more.

Lynn Reede: So this is Lynn. So I think the disconnect for many of us on the committee was that the way you're talking about an eMeasure was that we were still looking at it as if we looked at all of our measures for endorsement going through that same process and the specifications we're concerned.

So, is there a different way we should look at this than the way we did the day we met?

Elisa Munthali: Hi, this is Elisa from NQF. Yes, that's a great question. You should be looking at it the same way as you looked at it when you were here at NQF. So, it did pass evidence. It sounds like there are concerns about how the evidence is aligned with the specifications and that's a concern that would be something we would (strike) as a concern for other measures that are recommended for endorsement.

Again, this is one that is you are looking to approve for trial use, so we are going to see how the developer is able to test it, not endorse, but just wanted to provide that clarification.

(Multiple Speakers)

Helen Burstin: Looked at in the same way in terms of an expectation that's being reliability and validity now. We would see that later...

Elisa Munthali: Right.

Helen Burstin: ...when the eMeasure comes back for final review. It would not be endorsed. It would only be approved for trial use.

Lee Fleisher: So, Helen and Elisa, so the disconnect between the committee's concerns and the way the developer interpreted it were -- are we aligned now or are we still misunderstanding, because that was the response that I saw in the memo was more about that we didn't understand. The committee did not understand the correct way to evaluate it.

Helen Burstin: Yes, Lee, its Helen. It's hard to know, it's probably just be useful not worry about what it is then versus what it is now. And really just -- the committee did vote on evidence so I think this one did pass evidence (under)...

Elisa Munthali: Yes.

Helen Burstin: ...the vote in front of me. And I think the issue the Joint Commission is raising is, was this measure really (held) to the right standard for an eMeasure review, for a measure coming in for trial use.

Elisa Munthali: Yes. And measure coming in for trial use must meet our evidence bar as well and it did. And so I think the committee is struggling with how the evidence is reflected in the specifications, you know, the denominator, the numerator, all of what we require for the specs.

So, I don't know if today's conversation has helped you further with that discussion.

Are there any other particular questions you have for the developer to help you alleviate some of the confusions?

Richard Dutton: This is Rick. I'll probably add to the confusion. But, my concern is with the validity of the measure. While I think the committee agrees that it would be great to avoid inappropriate transfusion, I think the experienced clinicians on the committee are also very concerned that this would be measuring the wrong thing.

In other words, the decision to transfuse a patient shouldn't be based on hemoglobin but rather on, are they actively bleeding, what's the velocity of the

bleeding, what are the other comorbidities of the patient, what is the risk from hemorrhage and so on.

So this has been my problem with this measure from the beginning is I don't think it's measuring the right thing.

Helen Burstin: So, this is Helen again. It sounds like it's more of an issue of, do you believe this is a valid indicator of quality, not that you're expecting to see reliability and validity testing at this time which would not be appropriate for trial use measure. Is that correct?

Male: Correct.

Richard Dutton: Correct.

Barbee Whitaker: This is Barbee Whitaker. I'm sorry, I was on -- listening online and didn't realize I had to call in to speak.

I actually think that this is an important thing to measure whether we decided that we want to implement it as a full quality measure later on after we looked at it. Because you can't -- there's a lot of evidence that indicates that transfusing it lower hemoglobin levels is better for the patient. And, if you measure it, you can start to look at outcomes. It doesn't necessarily mean that if you transfuse a different hemoglobin levels with different conditions, it's not a good thing. It's just -- the measuring of it is an important first step.

Barbara Levy: Well, this is...

Mark Jarrett: This is Mark.

Barbara Levy: ...Barbara, Barbara Levy. I think that in a non-acute situation, that may be true. But, in an acute situation, you're not going to stop and wait for hemoglobin in an actively bleeding patient or, for example, with the postpartum hemorrhage. We teach exactly the opposite that if you wait for the hemoglobin to drop that low, she's off the cliff already.

So, I think the specifications of the measure and the way that it's being reported is the problem. And I think it's got some clinical problems associated with it.

I agree in a chronic condition, where you have a chronic medicine patient or something that this may be perfectly appropriate, but then we have to specify the measure in a way that carves out and eliminates those patients that are acutely bleeding and hemorrhaging.

Barbee Whitaker: I thought that was part of our recommendation was that actively bleeding patients are postpartum patients where we recommended that it would be removed from the denominator and numerator.

Michelle Dardis: And this is Michelle Dardis from the Joint Commission. May I speak to that point real briefly?

We did consider the exclusion of the postpartum and perinatal patients. And it was the preference of our (test) based on the transfusion evidence to continue to include the perinatal population in this measure.

The active bleeding population is an exclusion that we considered and this comes down to an issue of the measure format. With an eCQM capture of a condition of active bleeding would require active bleeding is captured in a form like the problem on the problem list to structured and encoded clinical data. And we feel there is limited feasibility based on our feasibility assessment which is part of the approval for trial use preparation to collect that data.

So our intent was to initially capture a population of all patients and use the data we received in these files which includes diagnosis data to see how we can work closely, identify that population we seek to exclude, the traumatic population.

Frederick Grover: But there are -- again, I -- you know, I struggle with this because I understand where you're coming from on one hand, but on the other hand, it can be confusing and send the wrong messages perhaps without some of these other variables being added in. And that's what concerns me.

If I was going to do something like this, I would include the hemoglobin in there but I will include some of the other things that we've mentioned that can affect your judgment. And I think when you do something that's incomplete like this, it can lead to erroneous implications that can affect, in a negative way, patient care.

Allan Siperstein: Allan here. No, I agree. My major issue with this is that of the scientific acceptability and that the, you know, although trauma patients are excluded, the -- there are a huge number of patients that are transfused acutely in the operating room or preoperatively the (redundancy), you know, life-saving measure to the hemodynamic change or active bleeding. And, you know, those patients, you know, really doesn't make sense in the context of this measure.

And the second scientific issue, I think, I hinted at was that of the stratified reporting of results based on hemoglobin levels, really makes it very difficult to have any scientific interpretation of the results.

Barbee Whitaker: I think some of the recent researcher has been looking at it less than seven, seven to eight, eight to nine, so -- or eight and above and 10 and above. So, that might be the reason why it's stratified as such.

Frederick Grover: That's true.

Christy Skipper: OK, if there are no other comments or discussion, we will move to vote on the measure specification. The voting -- that the specifications are consistent with the evidence, one high, two moderate, three low, four insufficient.

Female: And, Christy, just making sure that you saw the note because that's going to affect our count.

Christy Skipper: I did see, yes.

Female: Thank you.

Female: But we still have quorum.

(Off mic)

Christy Skipper: OK, and just waiting on two more votes.

Lynn Reede: Hi, Lynn Reede. I keep voting but nothing seems to happen to the total.

Christy Skipper: If you're online, you could chat your vote in and send it to NQF Leader Participants or if you're...

Lynn Reede: Thank you.

Christy Skipper: ...comfortable voicing it, you may do that as well. Just state your name.

Lynn Reede: Lynn Reede, low.

Christy Skipper: So, voting is closed, 0 percent high, 3 votes moderate, 12 -- or 12 votes low, 1 vote insufficient. This measure does not pass on measure specification, so we will stop here and turn our discussion to our review of the discussion of member and public comments that came in during the commenting period.

Female: OK, so next on our agenda, we did theme our comments into two different themes. We received several comments that were in support of the committee's recommendations. We won't be spending our time reviewing all of those comments now unless a committee member would like to specifically call out a particular comment.

You did receive all of the comments in the comment table, so let us know if there is anything that you would like to discuss there. Otherwise, we'll move

onto theme two, which is evaluation and discussion of the sociodemographic status trial period.

We received two comments, specifically on measures 1550 and 1551. The comments expressed concern related to the lack of rigor and robustness of the risk adjustment reviews and suggested that other SDS factors must be considered to understand the potential impact on a hospital's performance.

The commenter stated the conceptual model associated risk variables and results were not adequately discussed for these two measures. The committee did discuss this issue and ultimately the committee agreed not to recommend risk adjustment for SDS.

In addition to our NQF response that is in your memo, we did e-mail you on Friday a response from the developer CMS Yale CORE, their response to those comments.

And with that, we'd like to open it up to the committee to discuss these comments and potential response.

Lee Fleisher: So this is Lee. You know, some of the -- as we discussed there, I'll reiterate my question regarding the strength of the -- at least the variables that they looked at, this is one unlike other measures that has a conceptual framework for including SDS, the variable. I think one was 1.22 but they certainly were significant as the same level of a clinical variable.

They talked about deconstructing the influence of the hospital versus the patient. I'll be honest that beyond -- I don't fully understand what they were doing there, so I wouldn't mind having to better understand that, which was the rationale of not incorporating the SDS particularly given we would see both the SDS risk adjusted and the un-SDS risk adjusted in the measure which is what the -- what was (caught) for by the NQF Board.

So, I just will reiterate my previous comments that I continue to have some of those concerns.

A.J. Yates: This is Yates. And I would concur with what was just said by Lee. And I think it would be better to your trial of this if they were side by side, and you could see them and let the further stratification that was done stand for itself. And I'm not too sure that the burden of poverty is fully captured by just looking at the population.

The one part of the quintile as opposed to the entire -- in the higher quintiles, I think looking at that one part of the hospital's population doesn't capture the sense of what's going on in the entirety of the population that the hospital takes care of.

And again, I've made this analogy just because you take -- you look at the 20 percent of patients that might be impoverished, so it doesn't mean the rest of the people that are going to that hospital live in Beverly Hills.

And so, I think that there's an overall community effect that reflects the problems that exist when somebody lives in a tenement on the fifth floor with five sets of stairs, no elevator and has no car and public transportation is difficult and five -- you know, five bus exchanges to get to the hospital during the work day. I think that there is a difference between that and somebody that might live in suburban California with a car. And when you look at them and talk to them about the second floor, they look at you funny.

So I think that that -- we have to continue to look at these. I think that some of the risk factors chosen such as dual eligibility and ethnic status are surrogates for what is a, you know, significant burden of poverty for some hospitals that aren't necessarily bad hospitals. And I still disagree with bad hospital arguments. And I wait more clear evidence from Yale CORE and published literature to make that point. That's my two cents.

Frederick Grover: Where are you, page wise, in our report here?

A.J. Yates: This was in -- they're on the -- they're on comments from the outside talking about the -- what data points were used for the (assessed) and the risk

adjustment. And, there was a reply by Yale CORE making an argument that they used the most valid ones that were available. And I will give them that they did do that.

Katie Streeter: This is Katie. I'd also like to give our developer, representatives an opportunity to respond if they'd like to.

Karen Dorsey: Hi, this is Karen Dorsey from Yale CORE. Thanks for having the opportunity to respond.

You know, so, I would just start by reiterating that, you know, we agree that there are some important relationships between socioeconomic status and health outcomes undoubtedly. But, you know, what we'd contempt to do with our analysis was both lay out some of those conceptual relationships but then conduct empiric analysis to help understand whether the specific approach of patient level risk adjustment using the data that are available as you all just mentioned, you know, for use in the national measure like this would make -- have any impact on the measure results and would alleviate. I think, what the latent concern here, which is that hospitals that serve a disproportionately large population of patients who are lowest yet may be disadvantaged in terms of the measure results.

So that was really what we sought to do with the empiric analysis. And when we did that, you know, we were able to look at the results, hospital level result, side by side with and without risk adjustment. We found that risk adjustment had a negligible impact on hospital level results that just did not move results for individual hospitals very much at all.

And so that's really where we came down to. We tried to supplement that a little bit because as you all mentioned, there was some statistical relationship between some of the SDS variables and the outcomes to try to break down the proportion of that relationship that was due to the hospital level effect and the patient level effect showing that, you know, being more likely to go to a poor quality hospital was carrying a lot of that effect for measures.

So, just another sort of piece of evidence demonstrating that simply adding this to the patient level risk adjuster is not the most appropriate way to address, I think, the latent problem that you guys have discussed and we've articulated together in our discussion.

So, I appreciate the committee acknowledging. We did use the data elements that are available and valid right now. I think the National Academy of Medicine report that just came out in October, if there was any doubt about that really providing some independent evidence that that was true that we used to -- what was available at this time. And, of course, we remain committed to continuing to look into these and explore these analyses as your data sets become available.

A.J. Yates: If I could just have a point of clarification, you just said that you found no evidence to suggest that it helps with the risk adjustment. But correct me if I'm wrong, when this was presented and before you did the hospital versus patient stratification, the hospitals at the highest quintile penetrance of the three variables that were examined versus the hospitals with the lowest quintile, there was a definite difference. Am I not incorrect that that was found?

Karen Dorsey: Right, so there is an association between outcome rates and proportion of, for example, patients with lower SDS indicators. But what I'm -- what I was trying to say was that when we put it into the risk model and we look at the hospital level results of the risk standardized outcome rates with SDS variables in the model and without, we do not see movement in hospital results. Right, that's the -- it does not affect the final output or the measure calculation which is the risk standardized rate.

And you just reminded me that I did not mention the issue of -- sorry, I just lost my train of thought. I apologize.

Lee Fleisher: Can I follow up on that? So -- because you said it has a negligible effect and then you said it has no effect, and I'd love to -- and you said previously, and this is Lee again, just to clarify. But you didn't think it addressed the issue at

hand which is different than the model. So, I just want to understand, is it no effect, nothing moved...

Karen Dorsey: Right, sorry. Sorry. There -- you know, we saw a very small proportional change in the outcome rate. So there's some change, but it is very small. And if I don't -- I should be precise with my language, right? There are some difference in outcome rates among hospitals when you have the variable in or out, but that change is very, very small, too small to reasonably affect where hospitals fall in terms of their, you know, the bucket analysis that we use to check the (right) hospital performance or really pain adjustment.

Lee Fleisher: And just for clarification because you said it doesn't address the issue or concern. So is that a statistical issue or is that a -- what's the basis for that comment?

Karen Dorsey: Right, sorry. So, if you don't -- if in fact, including SDS variables in the risk model, does not alter hospital performance in particular, if we're concerned that hospital puts the larger proportion of these patients, right, are disadvantaged. If you don't see movement when you risk adjust, then there -- that is not a mechanism to alleviate that any potential bias, right, towards those hospitals. That's what I mean. Is that clear?

Frederick Grover: So, this is Fred, Karen, just ask you a question, does that mean then that for -- I mean, I think those of us who take care patients both in the private or university setting versus the city county, we -- you know, we think we see a difference in these patients. But what I'm (leading) up to is that difference may be captured in their regular clinical risk adjustment model such as...

Karen Dorsey: Absolutely. Oh, sorry.

Frederick Grover: ...more advanced disease, frailty, various risk factors. What we're always worried particularly about is what do you do when you -- what's the best way to discharge these people and ensure that they have reasonable follow up, and some of them are homeless, so all of those types of things.

But we probably -- I guess what you're going to -- what you're finding is that the hospitals, the safety net hospitals, have programs that probably address that issue, maybe that corrects for that even.

Karen Dorsey: Right, I mean, we can't see that, right, with the data that we have. But you're absolutely right that when we put in clinical risk adjuster so, you know, we built this analysis and we presented to you all so that you all could see the sort of bivariate relationship between the SDS variables and the outcome.

And then we added in the rest of the clinical risk variables that we use for risk adjustment. And you could see that they were responsible for the (alliance) here at the association, right? Once you put those clinical risk variables in the independent association with the SDS variables diminished a great deal.

Frederick Grover: OK. Well, that's an important -- you know, that's an important thing to, you know, objectively find. Because we've always been concerned that maybe the safety net hospitals are (taking) a real big disadvantage, but it sounds like if they are, they're making the adjustments. And most of them are making the adjustments they need to make.

Karen Dorsey: OK.

Barbara Levy: Yes, Barbara and Fred, this is Barbara. I think that -- I think the clinical adjustments is one thing, but I also think that when we talk about available data particularly with respect to resource use, the available data is what Medicare or the payers will pay, it's not what something actually caused these hospitals to do the work that they have to do with these underserved populations.

And I think they may be significantly disadvantaged in areas where we can't see the data because we don't really see what it costs to take care of them ...

Frederick Grover: Yes, yes.

Barbara Levy: ...and to follow up. And I think that's a really an important issue for us to consider. I'm not sure how we deal with it, but we are faced with this periodically and added that the available data may be inadequate. And, to create a measure with inadequate data may end up generating a measure that's not really valid. So, I just think that's something we need to consider as we move forward.

Lee Fleisher: So, this is Lee again. What would a statistically significant -- what would the Yale group feel would be, you know, an odds ratio 1.22, would be significant enough to incorporate it into the model so that I just understand?

Karen Dorsey: So we do not have sort of an odds ratio threshold for what makes a variable significant enough to be in a model. The way that we try to approach this was to say, you know, is there a conceptual relationship and then to sort of methodologically walk through certain steps to determine whether risk adjustment was appropriate or not based on the empirical evidence.

So, we started with a conceptual model that said, you know, there very likely may be some different burden of disease, there may be some extra residual effect that has to do with broad or societal issues that may or may not be with any control of the hospitals, and we wanted to look at how that played out in the analysis.

When we -- as I just stated, when we added those variables to -- and when we look at the risk variables and their relationship with the outcome, we did see a statistically significant outcome. When we add the rest of the clinical information, we see that that's attenuated to a large extent by the burden of disease that patients have. There is some residual statistically significant impact.

The next question that we were asked to look at was whether that -- whether or not we could tease apart the patient level and hospital level components, so we designed and performed an analysis to look at the hospital and patient level component. The argument there being that hospital level effects are not things that we tend to risk adjust for. We try to risk adjust for things that are

measured at the patient level, has a little bit of complexity even with that, with socioeconomic variables, but I won't get too much into that.

But what we found was that unlike many of the clinical variables that we include in the model, the hospital level effects really predominates for a lot of the socioeconomic variables or the rates variable that we were using, which demonstrates that these types of variables tend to operate more at the hospital level, meaning, it's an effect of patients going to lower quality hospitals rather than an intrinsic patient level characteristic exerting effect on the outcome.

And for that reason, we recommended that these not be used, you know. In addition to that, we also exert incredibly small differences on risk standardized outcome rate. So in the risk-benefit ratio here, you're not gaining much any real resolution of the problem when -- if the underlined problem is you're mischaracterizing hospital -- the hospitals are being characterized incredibly consistently with or without risk adjustment. And you're losing on the side of potentially including a hospital level of that with what's predominantly hospital level effect in the risk model which is not something that we want to risk adjust for.

So that was sort of the evidence and the rationale that we presented to the committee when we met this summer.

Lee Fleisher: Thank you.

A.J. Yates: I'm just looking at what's on the screen right now and under decomposition of analysis. And it found that, again, the dual eligible and low HER2 SDS Index scores were significantly associated with hip and knee complications. And so -- and then it was argued that this was hospital effect.

So in fact, there was something found when you did put it back into the model. I would argue that if -- that it's -- again, and I was told at the meeting that this was done but it hasn't been presented, that if you were to look at the whole spectrum of index scores across the entirety of the hospital versus other hospitals that there -- that you may capture the true nature of the community

that that hospital serves and that it's not that -- if it doesn't shut off at index of 42.7 that the rest of the patients that live in that community or go to that hospital may only be at 42.8 or 42.9. You don't know if you don't look at the entirety of the spectrum. And that would be one observation.

The other thing is, is that the model was exist -- the risk model that exists has a c-statistic of 0.65. So throwing these factors back into that model, is it possible that because the c-statistic is low and it's having a hard time that some of the factors that we're looking at get lost in there? I mean, is it part of the random walk of all the other complications that may or may not be associated or comorbidities may be associated with race or poverty?

So, I'm just (trying to) two things out to think of rather than, you know, just looking at the quintiles 42.7 below, but that it ought to be looked at our continuum. And the second thing is with the c-statistic being low for the model at large that we may be -- we still may be losing something that, in our literature, is arguably better supported as being an issue.

Karen Dorsey: Right, sorry, so, that's -- I lost my train of thought earlier and that's exactly what I wanted to speak to, which is the question about the treatment of these as continuous variables.

So, one of the reasons that we created dichotomous variables was so that we could really show the committee sort of apples to apples comparisons, right, our risk variables are yes-no variable to clinical risk variables. And so, we were really able to demonstrate direct comparisons with the other risk variables by dichotomizing variable this way.

However, of course, we did test in for several of our readmission measures, the use of the AHRQ SDS Index was a continuous variable. It made no difference, if it had, we would have, you know, reconsidered the benefit of the ease of presentation of using it as a dichotomous variable, but it did not make any difference to use it as continuous variable.

A.J. Yates: Yes...

Karen Dorsey: Which is why we made the sort of convenient choice of showing it the way we show the rest of our variables.

A.J. Yates: I understand, that's why I was done that way, it's just -- we didn't see this -- we didn't see the actual numbers of point given.

Karen Dorsey: I know your second point out. Let me just follow up by saying that, you know, as I said previously, we are, you know, very much committed to continuing to investigate in particular as newer and richer data sources that better capture concepts around socioeconomic status are available certainly, potentially things that can be more predictive and more appropriate for use.

A.J. Yates: I appreciate that because the literature and at least the orthopedic literature shows that there's a performance gap in that regard. And the more dialogue on this, the better.

Frederick Grover: So where do we go from here, do you think?

Female: So, our question to you as the committee is, do you find the developer's response acceptable or do you -- would you like to request additional information and further discussion?

A.J. Yates: This is Yates. Since I'm probably the loudest outspoken person on this, I would say that their response has been adequate -- more than adequate. They answered the specific concerns. And as long as there is, you know, recognition that we may not have the answer yet and that this is not a -- this is a two-year experiment by NQF and maybe the next go-around for -- next go-around of endorsements next year will somebody will show us something that this is the key.

So, I think we keep pushing ahead and asking the developers to do their best.

Lee Fleisher: So, can we -- for full transparencies since this is a major issue, would the best thing to do is to get -- does everybody have the full Yale response?

Frederick Grover: You know, I'm not totally sure I do.

(Off mic)

Frederick Grover: Refer me exactly to these documents where I would be sure I've read the whole thing, Lee.

Lee Fleisher: Christy?

Christy Skipper: Yes. So, I apologize. Their response came in on Friday afternoon and we did send that all to you.

Female: Does the committee needs more time to review?

Lee Fleisher: I'm just wondering and I'm not as -- you know, given the importance of this, is there a value in just ensuring -- if everyone's comfortable not reopening it, I'm totally comfortable with that. I just want to -- I think transparency as Yale has tried to be -- as Yale is, we should be transparent. So, should we have the -- if anybody would like to reopen validity, we can, but otherwise, just give people a few days, Christy, what do you -- or Helen or Marcia or...

Helen Burstin: No, I mean, the real question is, you know, if people feel like this discussion, the response from Yale or the -- particularly the responses you hear from Karen today and the back and forth with Dr. Yates, is that something you feel comfortable that we should just continue to move this measure through as is, has it raised any concerns that you would want additional time to view the materials from Yale, we can sort of follow up on e-mail with the committee.

But again, it's really -- I'm not sure re-voting per se that changes anything at this point. I think it's more a question of, if you feel like the questions are answered adequately by Yale. And do you want any, you know, time to consider them further since they came quite late.

A.J. Yates: Again, I've read the response and I thought the response was spot on in terms of pointing out that they use the most valid criteria to look at. And quite frankly, out of all of the different groups to look for re-endorsement or endorsement on this round, I think Yale did the hardest -- made the greatest effort to actually answer the question. And I think at this point in time, we don't need to reopen -- I think we should agree that it's OK to leave the answer as-is.

Christy Skipper: OK. So what we're hearing is that you accept the response?

Male: Yes.

Female: OK.

Male: Yes.

Christy Skipper: OK.

Helen Burstin: And what's really (reflect) this discussion in (memo) when it goes to CSAC.

Frederick Grover: Yes, I'm comfortable with that as well.

Helen Burstin: Great. Thanks, Fred.

Female: OK.

Christy Skipper: OK, operator, could you open the line for any public or member comments?

Operator: If you would like to make a public comment, please press star one.

And there are no public comments at this time.

Christy Skipper: OK. Thank you. So, at this point, our next steps are, the team will be working to update the draft report based on your discussions today. And then,

all of your recommendations will be moved to the feedback for discussion on December 13th and then to the Board for ratification in January.

So, if there are no other -- if there are any questions, you can ask them now, if not, we will end the call here today.

Helen Burstin: And I'll just point out -- this is Helen, just one last thing. I do still want to get a group of you together to help talk about the issues of -- concerns about how we would look for unintended consequences and concerns about cherry-picking.

So, perhaps, we'll have Christy to send a follow up e-mail after this call asking for volunteers, and we'll just schedule a call to follow. And as you heard even from some of the developers at the meeting, I think it was the Yale, in putting (it direct) should be useful to hear from the committee of what kinds of unintended consequences or how would you even look for some of those concerns related to some of these measures. So, more on that to follow.

Thanks, everybody.

Frederick Grover: Thank you, Helen.

Christy Skipper: All right, have a good afternoon. Bye-bye.

Allan Siperstein: Great. Thank you very much.

Female: Bye.

Operator: This concludes our call. You may now disconnect.

END