NATIONAL QUALITY FORUM

+ + + + +

IN-PERSON MEETING
PATIENT SAFETY STANDING COMMITTEE

+ + + + +

THURSDAY
JUNE 18, 2015

+ + + + +

The Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:30 a.m., Ed Septimus and Iona Thraen, Co-Chairs, presiding.

PRESENT:

ED SEPTIMUS, MD, Co-Chair, Medical Director Infection Prevention and Epidemiology HCA and Professor of Internal Medicine Texas A&M Health Science Center College of Medicine, Hospital Corporation of America
IONA THRAEN, PhD, ACSW, Co-Chair, Patient Safety Director, Utah Department of Health
JASON ADELMAN, MD, MS, Patient Safety Officer, Montefiore Medical Center
CHARLOTTE ALEXANDER, MD, Orthopedic Hand Surgeon, Memorial Hermann Health System
KIMBERLY APPLEGATE, MD, MS, FACR, Radiologist/Pediatric Radiologist & Director of Practice Quality Improvement in Radiology, Emory University *
LAURA ARDIZZONE, BSN, MS, DNP, CRNA, Chief Nurse Anesthetist, Memorial Sloan Kettering Cancer Center
RICHARD BRILLI, MD, FAAP, FCCM, Chief Medical Officer, Administration, Nationwide Children's Hospital

CHRISTOPHER COOK, PharmD, PhD, Director, Quality and Performance Measurement Strategy, GlaxoSmithKline
MELISSA DANFORTH, BA, Senior Director of Hospital Ratings, The Leapfrog Group
THERESA EDELSTEIN, MPH, LNHA, Vice President of Post-Acute Care Policy, New Jersey Hospital Association
LILLEE GELINAS, MSN, RN, FAAN, System Vice President & Chief Nursing Officer, CHRISTUS Health
STEPHEN LAWLESS, MD, MBA, FAAP, FCCM, Vice President, Quality and Safety, Nemours
LISA MCGIFFERT, Project Director, Safe Patient Office, Consumers Union
SUSAN MOFFATT-BRUCE, MD, BSc, PhD, MBOE, FACS, FRCP(c), Chief Quality and Patient Safety Officer, The Ohio State University
ANN O'BRIEN, RN, MSN, CPHIMS, National Director of Clinical Informatics, Kaiser Permanente
PATRICIA QUIGLEY, PhD, MPH, ARNP, CRRN, FAAN, FAANP, Associate Director, VISN 8 Patient Safety Center, Department of Veterans Affairs
VICTORIA L. RICH, PhD, RN, FAAN, Hospital of the University of Pennsylvania
JOSHUA RISING, MD, MPH, Director, Medical Devices, The Pew Charitable Trusts
MICHELLE SCHREIBER, MD, SVP Clinical Transformation and Associate Chief Quality Officer, Henry Ford Health System
LESLIE SCHULTZ, PhD, RN, NEA-BC, CPHQ, Clinical Consultant, Premier, Inc.
LYNDA SMIRZ, MD, MBA, Chief Medical Officer and Vice President of Quality, Universal Health Systems of Delaware
TRACY WANG, MPH, Public Health Program Director, WellPoint, Inc.
KENDALL WEBB, MD, FACEP, Associate Chief Medical Information Officer, University of Florida Health Systems

YANLING YU, PhD, Physical Oceanographer and Patient Safety Advocate, Washington Advocate for Patient Safety

NQF STAFF:
HELEN BURSTIN, MD, MPH, Chief Scientific Officer
ELISA MUNTHALI, MPH, Vice President, Quality Measurement
ANDREW ANDERSON, Project Manager
JASON GOLDWATER, MA, MPA, Senior Director
LAURA IBRAGIMOVA, Project Analyst
JESSE PINES, MD, Senior Director
SUZANNE THEBERGE, MPH, Senior Project Manager

ALSO PRESENT:
MARY BARTON, MD, MPP, National Committee for Quality Assurance
SVEN BERG, MD, MPH, CPE, West Virginia Medical Institute
NATHANIEL BREG, BA, RTI International
ALYSSA CRAWFORD, Mathematica Policy Research
THOMAS W. CROGHAN, MD, Mathematica Policy Research
CYNTHIA CULLEN, MBA, Mathematica Policy Research
RICHARD DUTTON, MD, MBA, American Society for Anesthesiologists *
ERIN GIOVANNETTI, PhD, National Committee for Quality Assurance
DANIEL GREEN, MD, Centers for Medicare and Medicaid Services *
NICOLE KEANE, MSN, RN, Abt Associates *
QINGHUA LI, PhD, RTI International
EUGENE NUCCIO, PhD, University of Colorado Denver
DANIEL POLLOCK, MD, Centers for Disease Control and Prevention *
ANGELA RICHARD, PhD, RN, University of Colorado Denver *
LAURA SMITH, PhD, RTI International
WILLIAM SOUTHERN, MD, MS, Albert Einstein College of Medicine, Montefiore Medical Center *
* present by teleconference

TABLE OF CONTENTS

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

8:26 a.m.

CO-CHAIR SEPTIMUS: Good morning, everyone. Good morning. Good morning. Who's on the phone first? Who's on the phone?

MS. O'BRIEN: Good morning, this is Ann O'Brien from Kaiser Permanente.

CO-CHAIR SEPTIMUS: Okay, who else? Is that it? Any Patient Safety Committee members on the phone this morning? I guess not yet. Okay, and we really appreciate those of you who were in there almost all day yesterday on the phone; thank you very much. It's extremely difficult. So I hope everybody had a good dinner; I had to apologize to the group, there was no Septimus wine on the wine menu last night, which I know disappointed many of you. We had really a very productive day yesterday. Just to sort of recap, we tabled two measures on the failure to rescue; we'll take those up on our call post-meeting. We put on reserve status 0038, and as you know for PSI 90, we only

achieved a 58 percent, so short of the 60 percent consensus, so as you know that's in a gray zone. We're going to wait for public comment and we may come back and discuss this after public comments are in.

So as you know because of yesterday, we made some slight changes in the schedule. That was emailed out, but just to go over, this morning before the break--we hope we can get to a break--we're taking up 2726, which is Prevention of Central Venous Catheter-Related Bloodstream Infections from the American Society of Anesthesia, then 2720 is a new measure, Antimicrobial Use Measure from the CDC, and then 2729, which is Timely Evaluation of High-Risk Individuals in the ED, and then 0687, Residents Who Were Physically Restrained in Long-Stay, and then 0689, Residents Who Lose Too Much Weight, also in Long Stay. Both of those, the developers are CMS. And then we hope we can take a break, and we'll tell you about the schedule after that.

Also, we're very excited, we're going

to take up our first eMeasures today, which is--I think--is it a first for this committee? For this committee, right, so it's a first for this committee, so we're very excited about that. And so to introduce us to the eMeasure evaluation--where's Jason--he's behind us. Jason, get up to the--we re-named you the measure developer Jason. So Jason will take us through that, and then we'll get right into the agenda, starting with 2726.

MR. GOLDWATER: Thank you very much, and good morning, everyone. I am sure that when you woke up this morning, you were bristling with excitement over learning about eMeasures. I can't blame you; I get excited about this every time I talk. It's all over my face, I'm sure. So my name is Jason Goldwater, I'm a Senior Director here at NQF, I've been here since January. I have a multitude of responsibilities, as does everyone that works in this organization, but my primary responsibilities are overseeing our Health Information Technology Group and

portfolio of work, and also overseeing our eMeasure development and approval process.

And eMeasures have somewhat really come into the forefront over the last several years. The passage of HITECH in 2009 initially was designed to serve as a catalyst to increase adoption of electronic health records throughout the United States, particularly in hospitals and what was defined as eligible providers, as well as devising a baseline to evaluate how EHRs could be used in the course of care, and that of course was meaningful use. And so this is not a discussion about the ins and outs and merits of meaningful use, but rather the part of meaningful use that dealt with what you will be discussing later on, which is the development and use of electronic clinical quality measures. eCQMs--is the short form for this--have been around for some time. It's not like it's a brand new occurrence, but certainly they have taken on much more life since 2009, and particularly over the last three years as EHR adoption has increased

substantially.

The reason why eMeasures have taken on so much more attention, have been given so much more attention, is because the availability of data, real clinical data to populate and test and use a measure, is much more available than it was in the past. Part of my career, which is long and diverse and incredibly strange, which I will spare you all of the details, but I did work at CMS for 10 years, when it was called HCFA. You all remember the good old days when it was--refer to the name as it should continue to be referred to as--well, that's my own opinion. So I worked for what was then OCSQ, which apparently has changed to CCSQ, during the Eighth Scope of Work for QIOs in which paper-based, traditional claims measures were still highly prevalent, but at that point in time, which was 2003, CMS was really trying to make a move into developing and utilizing eMeasures, and the first project they undertook, which some of you that are real historians of HCFA might remember was the

Doctor's Office Quality Improvement Technology Project, or DOQ-IT for short, in which they wanted to take advantage of the increase in EHR adoption to see if there was a way of automatically generating electronic clinical quality measures rather than through what was the traditional paper-based claims abstraction process, through the CDACs, at which time when I joined OCSQ, there were four. There are now--I think there's one, and I'm not even sure if that one is still around.

DOQ-IT did not succeed as well as I think everyone had hoped, for a couple of reasons. One is there was not widespread EHR adoption. There was only maybe a 21 percent adoption rate at that point in time. So really, culturally, the country wasn't ready for this. And secondly, they found out that trying to generate electronic clinical quality measures from data other than claims was an incredibly trying and very difficult process, something that still is pervasive to this very day. So what I

wanted to talk about in, I guess the next 10 minutes, is sort of what you need to look for when you're considering eMeasures, and what NQF does when an eMeasure arrives, and what we do when we're performing our evaluation, and how that is going to differ from what you're going to be asked to do. Under no circumstances would you ever be asked, at least unless something is going--somebody is going to pop a surprise in the next two minutes, where you would be asked to look at formats or codes or markups or value sets. Those are things that we take care of when we do the evaluation, but really when a committee is evaluating an eMeasure, it's not that much different from the way you look at a traditional measure, there's just a few things that need to be considered.

When an eMeasure arrives to NQF, we do an assessment very similar to the way we would do it with a normal, traditional-based measure. We look at its evidence, we look at the scientific acceptability of the measure as we would with

anyone, we look at its reliability, we look at its feasibility, we look at its validity. There's a feasibility scorecard that every eMeasure developer must submit to let us know how the eMeasure, when it was tested--now remember, when an eMeasure is tested, it's not tested in the same way a claims measure is; it does need to be tested on an EHR system. In the past, it used to be three or more EHR systems.

That is difficult, not just simply because of trying to find three diverse EHR systems to test, but also actually asking the providers or the hospitals for a specific period of time to generate the data necessary to appropriately test the measure is a time-consuming process.

Our policy has changed over that time that we've gone from three to just more than one, and allowing measure developers to use at least more than one system to try to test their measures so we can get the results to determine whether it can feasibly be done within an EHR in

the normal course of business.

We also look at the formats, and what I mean by formats, without getting into a lot of detail, is when information is transmitted from one system to another, in order for it to maintain its integrity and meaning, it has to be formatted in a very specific way. We use something known as the Health Quality Measures Format, which, if I really were to get into the higher level of this, is just like a basic webpage, in all honesty. It's marked up in very much the same way, it's marked up in a mark-up language that if--for those of you that can go way back to the days of devising, creating web pages; have any of you ever done that? None of you have done that, all right. So fine.

Back in the glory days when the internet was first becoming this ubiquitous term that we all know and love, you actually had to design web pages in what was known as hard-coding, which is you actually had to type in the codes of how the page was supposed to look. And

I often joke with people that when you had to create moving graphics--which nowadays is ridiculously easy--back then you actually had to program how you wanted the graphic to go from left to right or top to bottom, and you could never get it right, and you would spend an hour typing in things such as inches or centimeters just to get it correct. That's very similar to what you have to do with an eMeasure, but that's not something you would ever have to look at; that's what we look at to make sure that it is coded and formatted appropriately, and that it can move from one system to another with its meaning and integrity still intact.

We also look at the way it would be outputted, so CMS requires it to be reported on what they call a Quality Document Reporting Architecture, which allows reporting on individual patients or population of patients. We check the formats to make sure there will be a seamless transition from one to the next. Again, not something you have to worry about. We also

look to make sure that the right measure artifacts are in place, that the right files are there, that we have the electronic information that we need, that it's complete because that document is a complete and comprehensive electronic measure. And then it gets to you. So then what do you have to do?

There's really four things you really have to think about when you're looking at an eMeasure. In terms of evaluating the numerator or denominator or an exemption or exclusion, nothing changes from evaluating an eMeasure as it would be over a traditional measure. In examining its reliability, its feasibility, its validity, again nothing truly changes, and we will have written our comments and our concerns about what we think about those topics when we are evaluating the measure for your consideration. Other things you really need to consider are the following: the first is, is the measure good enough? And I think that's kind of a basic question, which is yes it's an eMeasure,

and yes it's contained within an EHR, but in your opinion is this measure good enough? If you're going to have to go through the work of actually getting the data in a structured format and exporting that data out of an EHR, is the measure viable, is the measure necessary, will the measure actually make improvements in quality in your mind, and is robust and good enough that it should continue?

The second is can the data that's needed to populate this measure actually be obtained within the system it's supposed to come from? Do you think the data is there in the--now all of you work in these clinical environments; you're going to have a better assessment and understanding of this than anyone--do you think that that data is available and obtainable and can be taken out of a system so that an eMeasure could be generated and utilized and evaluated? If you really think it's going to be highly problematic, then that's a concern. But you have to look at it not in terms of is it on a claim,

which is somewhat easier, but can it be found within the very system it's supposed to come from.

Thirdly, and I think this is important, there's always this talk about EHRs improving the efficiency of care, and that's true. There's a lot of efficiencies that come with it, particularly in the area of patient safety. I mean, there are a lot of--there's significant functionality within EHRs to improve patient safety. How successful that is is an entirely different discussion, but in order to ensure that the data is there in the system that can be used for the measure, it does have to be inputted. It doesn't magically just somehow generate itself within the system.

So in your mind, in the normal course of providing care to patients, is that something that would be entered in the normal course of a workflow, or is it something that would be so overly burdensome for some--a nurse or for a physician to do that it would be almost

impractical to be considering this as a measure because it would be impossible for the data to be entered and even harder for the data to be extracted? Those are the things you have to look for, and I think workflow sometimes is understated a bit, that it's not considered in the evaluation of an eMeasure, and I think that's important to consider because, again, in order to get the data out, the data has to get in, and if you really think that this is too complex for a data to actually be put into a system by a provider or by a nurse, then it would be virtually impossible to be actually generating any type of eMeasure.

And then the last one is do you think this eMeasure would be sustainable over time, which is in the course, you know the--I remember when the AMI measure, aspirin on arrival, like this is the oldest measure in time. Like I remember when I was significantly younger--in NQF years, I'm like ancient really, I mean I'm like one of the oldest people in this company really.

Seriously. It's depressing, very depressing, but I remember AMI, aspirin on arrival, being one of the very first measures around, and that came out because then, when it was released, the science indicated that if you give aspirin on arrival, somebody's having a cardiac event, it makes a significant difference. And of course over time, that has become so widely used and now it's common practice. It's been a sustainable measure over time, and even as they transition from traditional claims-based measures into electronic measures, that's still being used.

It's very easy to find the data within the EHR that somebody has presented with a cardiac event, indicating AMI, and they were prescribed aspirin on arrival, and here's how much aspirin they were prescribed. And for those of us that are real nerds about this, we can tell you the values and how it was coded and how to evaluate it, but it's very common. That's something to consider as well; do you believe if this measure is used, and that if the data is

going to be available within the system, and the data could be entered in the system in the normal course of workflow, do you also believe over time this measure would be sustainable, it would continue to make improvements in quality over time, or after a year, is this going to be so unbelievably burdensome to everyone, even if it has the best of intent, that it would no longer be used, in which case it's something to be considered for a possible endorsement, because that's--if you're only going to have a measure for a short period of time, do you really want to continue to engage in moving that measure forward?

So those are really the four things to be considering in the context of evaluating an eMeasure. Again, not overly technical; I don't think we've ever asked our committees to be specialties in the technical expertise of coding or formatting, but really it's more policy-level use questions, usability questions on how the data is used and how the data can be extracted to

create an eMeasure. Yes? Sure.

CO-CHAIR THRAEN: Sorry, so in your analysis, in your process, do you--since you're only--you changed your rule from more than one, okay, and there are like many three dominant players in the market right now, and each of them are using different standards-based, you know, how do you account for the reliability of the measure capturing what it is it's supposed to be capturing? How do you analyze that?

MR. GOLDWATER: Right, so that's an excellent question, and that was a big concern when we switched the rule, because there are--so there are 175 EHR vendors at the moment, but there's only about five of them that have roughly 90 percent of the market; Epic, Allscripts, and Cerner are the ones that have roughly 80 percent of the market. So if we were going to go to more than one, are we just going to have people testing out of Epic? So there's two--I don't want to say there's answers, but there's two sort of things to consider. The first is that even

though Epic may have 100 implementations nationwide, each implementation of Epic is different; it is not the same. They do have a common base system, which is what they sell, but they will change the base system and they will add options based upon what the hospital or the ambulatory provider needs. So there are different variations of the same system.

It is true that they do use some different standards, but really the standards that they use that are varying are the way the information is transmitted, and that's a separate discussion because Epic has its own proprietary standards in which information is transmitted, usually within a hospital or between Epic systems. However, the way the information is coded and the way those codes reflect a clinical content are similar whether it is Epic or whether it is Cerner or whether it is Allscripts. That's correct, at the data element level, it's very similar. So problems are SNOWMED codes, diagnoses are ICD-9, soon to be ICD-10 much to

the chagrin of some people, and then eventually, if you're ready for this, potentially ICD-11. Outpatient ambulatory codes are still CPT or HCPCS, laboratory codes are still LOINC, medications are still RxNorm.

So regardless of the diversity of systems, the coding at the data element level is still the same, which makes it easier to populate and understand the value sets within the EHR system, even though it brings up some issues of, you know, when we do the evaluation of the testing, we're really going to have to look at the--so if somebody does three tests, and they're all Epic, then our analysis has to be, and what we have to present to you is how the systems differed. Yes?

MS. DANFORTH: I just have two questions. When the eMeasures come to you, if there's already been a measure that's not an eMeasure, is there any kind of analysis or comparison? I'm saying when there is, do you do--yes, so when there is. So for example, like in

the IPPS proposed changes, there's a handful of measures that are going from paper measures to eMeasures.

MR. GOLDWATER: Right.

MS. DANFORTH: Is there--and the way that CMS worded it, it sort of implied that there's actually a difference in the rates produced from the eMeasures versus the paper measures?

MR. GOLDWATER: That's correct.

MS. DANFORTH: So in evaluating the measure, is that part of the discussion, like what those changes are and what the significance of those changes are?

MR. GOLDWATER: Yes, well not completely, but in our process or our policy is, if you're going to move from a traditional claims-based measure to an eMeasure, the eMeasure is considered new, and--because it is a new measure. Now that has brought up some policy questions we're still discussing, as Helen is well aware, and probably those conversations will

continue on for a while. But we do evaluate it as an eMeasure, and we have to look at the same things we would look at for any eMeasure. Now granted, there's a past precedent to base that on, but it is considered an eMeasure and has to be evaluated that way.

DR. BURSTIN: Just building on part of your question though, I think the other piece of this, and we'll see, is we don't actually know whether in fact the rates will be comparable, and it's actually I think an important piece of work we'd like to engage in going forward because I think it may truly change some baselines in ways that change trend lines, et cetera, which is why we feel like very strongly we need to really understand what those differences are. It's a great question.

CO-CHAIR SEPTIMUS: Excellent.

MS. DANFORTH: Can I ask one more quick question? So yesterday we spent a lot of time talking about sort of the inadequacy and sort of known issues with claims-based measures.

Is anyone looking at potential issues with standardized documentation that are going to create issues in clinical documentation-based measures, and is there just any--will that kind of information be provided to us as well? I mean, I think yesterday there was a lot of conversations about well, you know, I think Jason said when we find this problem, we look in the claims and we saw that it was a coding problem, right? So everyone's looking at coders and what coders are doing, but the documentation that's going into the EHRs is extremely important; if it's not done in a standard, high-quality way, then these measures aren't going to be any more accurate than the claims-based ones. So is part of the work that you're doing sort of looking at that and developing sort of a list of known issues, then letting the committee members know about that?

MR. GOLDWATER: Do you want to explain the measure developer?

DR. BURSTIN: So one of the things

you'll see from what Jason and the staff will present to you on these eMeasures is an assessment of the things we require for eMeasures, one of which is called eMeasure feasibility; it's a score where they actually look to see the data elements you would need to do this measure: are they something you can actually find in an electronic record? It's still pretty early, and I think one of the reasons we shifted from saying three or more EHRs to more than one is it's really hard right now to find EHRs to test measures in, and we don't want it to be a rate-limiting step, but we want to be able to get them out there, but we recognize there's got to be a lot of testing.

MR. GOLDWATER: And one of the other problems that we also have to look at, and it's actually interestingly a project we're taking on independent of this, which is the way that it measures--the way the values are developed with an eMeasure. So value sets play a large part in eMeasure development, and there are varying value

sets for all different types of measures, and at times those value sets will overlap, will be redundant, and will not have a lot of meaning to the actual measure that they're--they will have very little relevance to the intent of the measure. So we're working now on a project to sort of harmonize and remove that variance, but that's something else that we consider as well in the feasibility testing, which is what value sets are you using and how are those measures coded, and is it reflective of the intent of the measure, as well.

CO-CHAIR SEPTIMUS: We only have time for one more question, because we're running right up to the next measure, so--

DR. LAWLESS: Yes, Steve Lawless. My question for you is, are there lessons learned from both the implementation of meaningful use and the level 3 that way that you're using to say, you know, lessons learned from trying to extract even simple stuff from an EMR, turning it into a measure that can be--that you're utilizing

or saying, aha, let's avoid these paths of this, and the other thing about EMR is it's very sensitive data and very specific, so how do you code it, you can miss a lot because you've missed a decimal place here versus somebody looking at what your intent was. So my more important question is the first one about lessons learned.

MR. GOLDWATER: So it's interesting that the final question is of course the easiest one to be answering. I can do it in 60 seconds-- it's an excellent, excellent question. And there's a lot of lessons to learn from meaningful use, and I think when it comes to eMeasures, the two most dominant ones have been is the information coded appropriately, so there are of course domains of terminology that ONC has really recommended to use, but are those being used correctly? Are those being used to adequately code diagnoses, procedures, medications, et cetera? Are the appropriate value sets being used that reflect the intent of the measure?

We have found that that has been

somewhat problematic, not as problematic as we expected, but then again the measures are coming in relatively slowly, so if we get 40 eMeasures, which would be great, and we get those in in the next couple of months, we may find those problems to be far more pervasive. And then secondly, you know, one of the problems has really been on the vendor side, which is, are they able to develop the system to be able to collect and report out the data based on the specifications that have been documented by CMS and that NQF has moved forward with?

And that has been a problem since meaningful use came out, and it's not just with eCQMs, it's just with everything. There are issues about whether the EHRs can function in the way that CMS would like, particularly with reporting quality measures, and will they get the information that is needed. That was a problem that existed when DOQ-IT was around; that problem hasn't exactly gone away, particularly when you start talking about measures that go outside the

traditional AMI, pneumonia, stroke, VTE, which have been around for a while. When you get into the behavioral health measures, the eye, ear, nose and throat measures, things that are newer, that becomes an issue because, do they have the ability to code and reflect that data adequately?

So those are things we have to really look at, which is why our eMeasure review process takes a little bit of time, because we have to really get into the nuances of those elements and the value sets to make sure they're done appropriately.

CO-CHAIR SEPTIMUS: I think we'll have opportunities as the eMeasures come forward to have some further discussion on this, and I hate to cut off conversation; this is the first time this committee has considered eMeasures. Before we get started with the first measure, is anybody--any other committee members joined the call?

DR. APPLEGATE: Yes, this is Kimberly Applegate; I'm non the call. Thank you.

CO-CHAIR SEPTIMUS: Okay, so we have two from yesterday, and did we lose anyone here? So Tracy. So we're down to 23 for voting, that's what I'm--22, that's what I'm counting. All right, so we have 22 for voting--

DR. BURSTIN: Josh just said he'd be in late; he has to drop his kids off--

CO-CHAIR SEPTIMUS: Okay so for now, it's going to be 22 for voting. Okay, so the first measure is 2726--thank you. 2726. So the first one is a measure that we looked at last year, Prevention of Central Venous Related Bloodstream Infections from the American Society of Anesthesiology; they'll present for the first couple of minutes, and then Dr. Alexander will be the discussant. Yes, I said 2726; didn't I say that?

CO-CHAIR THRAEN: You may have.

CO-CHAIR SEPTIMUS: Okay.

DR. DUTTON: Good morning, can you hear me?

CO-CHAIR SEPTIMUS: Yes, who--is that-

-

SPEAKER B: Yes, that's Dr. Dutton who's--he will be speaking with us today.

CO-CHAIR SEPTIMUS: Okay. Dr. Dutton, would you like to introduce yourself to the committee members, please?

DR. DUTTON: Sure. This is Dr. Richard Dutton, I am an anesthesiologist and the Chief Quality Officer of the American Society of Anesthesiologists; I also participate in some other NQF activities. Thank you for having us back to discuss this measure. As you've just heard from Dr. Septimus, we did present this last year. We had just taken over management of the measure from the AMA, from PCPI, and I have to admit the handoff wasn't good, so there were questions we were unable to answer last year. We've tweaked it a little bit this year, and we've completed the validity reliability testing with new data, which I think will help the presentation this time.

So you have the measure in front of

you. You've had the opportunity to review it. It's fairly simple. It calls for measurement of the use of preventive measures for preventing central line infection at the time the line is placed. As many of you know, depending on your hospital and your system, up to half of all the central lines in the hospital will be placed by anesthesiologists in the OR or ICU environment, so this is an important measure for our specialty. But very often we are putting lines in, but then not around or not managing the patient later when the complication occurs, and this is why we feel that a process measure is still appropriate for this activity, because the process and the outcomes are separated in time, and in our case, separated by professional service.

Successful compliance with this measure calls for using maximal sterile barrier precautions when placing the line, use of a cap, mask, sterile gown and gloves, full body drape on the patient, and if ultrasound is used, which is

recommended, sterile gel and sterile ultrasound probe cover. It applies to all patients regardless of age who have a central line placed, and it's a fairly broad definition of central line placement. Of course, it's hard to tell from the measure, which is full of CPT codes, but that covers both tunneled and untunneled central lines, so both temporary and longer term central lines, and it includes peripherally-inserted central lines, PICC lines. Current performance on the measure, looking at the data in our registry and now about four years of performance on this, about 60 to 70 percent of anesthesiologists report the measure when a central line is placed; we can see that in both Medicare data and our own registry data, so there's a significant gap in utilization of the measure and reporting it at all.

When it is reported as you might imagine, it is mostly successful, in the low 90s right now, but we also know that there are many practices and many physicians who achieve 100

percent or close to 100 percent performance on this measure. The connection with outcome has been very strong; AHRQ has recently published data on the rate of CLABSI central line infections in the United States, which have dropped precipitously since this measure has been in place, and since the use of sterile techniques when placing lines has been focused by anesthesiology and other specialties. So I'll stop there; I'm happy to take any questions or react to any comments from the committee.

CO-CHAIR SEPTIMUS: Thank you very much. Charlotte, do you want to take us through the measure, please?

DR. ALEXANDER: Certainly. You had an excellent introduction. You want me to go directly into evidence?

CO-CHAIR SEPTIMUS: Let's go to the evidence and go from there.

DR. ALEXANDER: As he stated, this is a process measure. Hospital-acquired infections are a common complication that leads to increased

cost and mortality; 51 percent occur in the ICU and central venous catheter is probably the largest risk factor. Catheter-related bloodstream infections commonly occur when the catheter becomes contaminated by microbes on the skin during insertion. Maximal barrier technique has been shown to be a cost-effective way to reduce these infections. There is a guideline which has 12 recommendations for sterile technique for insertion of these lines. There are also 14 studies that are high-grade, half of which are root cause and random control studies with large sample sizes. They're uniform in their evidence that maximal sterile barrier technique can decrease bloodstream infections from five to 35 percent consistently, and up to 65 percent in some cases.

CO-CHAIR SEPTIMUS: Any questions on the evidence? Just a great, great presentation. There's actually two components to prevention of central lines; one is on insertion, and early central line infections, the predominant

pathogenesis is colonization around the insertion site. Later in the course of central lines, it's really related to maintenance and contamination of the hubs, which would not be related to insertion. Just wanted to balance that in terms of the pathogenesis. But as is stated, they're looking at the so-called CDC guidelines and the level of evidence for insertion of the central line, and they're well outlined in your packet. So are there any other questions before we vote on the evidence? Yes, Steve.

DR. LAWLESS: I just--this is--but in the definitions and maybe the evidence, the evidence doesn't include the type of skin prep used?

CO-CHAIR SEPTIMUS: What skin prep are you talking about? It should be CHG alcohol.

DR. LAWLESS: That's what I'm saying. I see from the front sheet, it talks about, you know, the standard stuff, the cap, the mask and everything else and skin prep, but it doesn't

mention type of skin prep, or -- that's not part of the evidence?

DR. ALEXANDER: That's in the CDC guidelines.

CO-CHAIR SEPTIMUS: But you're correct, CHG alcohol is--

DR. LAWLESS: But it's included, that is included as part of the numerator here in terms of that --- that would be also being followed or looked at?

DR. DUTTON: Yes, and one of the small changes in our presentation this year is it--the last time around, they called for just use of chlorhexidine; it now includes alcohol, tinctured iodine or chlorhexidine as acceptable prep solution, based on the newest data.

CO-CHAIR SEPTIMUS: Yes, Lisa?

MS. MCGIFFERT: It looks--I want to-- is this just a measure for anesthesiologists, or is this a measure for anyone? And it also looks like --is it always going to be --it's not always going to be in a hospital setting; I'm just

trying to figure out who are we measuring here?

DR. DUTTON: The denominator codes for eligibility for this measure are CPT codes for insertion of central lines. So any providers who place central lines would be eligible to report this measure. The cyberstudy includes anesthesiologists, nurse anesthetists and others working in the OR; it includes surgeons, oncologists, intensive care physicians, or even nurse practitioners closing PICC lines.

CO-CHAIR SEPTIMUS: Are there any other--I don't see any, so let's go ahead and vote on the evidence.

MS. IBRAGIMOVA: So importance to measure and report 1A evidence structure process intermediate outcome, the votes are 1 high only eligible if QQC submitted, 2 moderate, 3 low, 4 insufficient evidence.

CO-CHAIR SEPTIMUS: Are we getting the ones on the phone through the WebEx? Okay, good. Two more.

DR. ALEXANDER: Susan stepped out of

the room.

CO-CHAIR SEPTIMUS: Okay so that's--that's going to be--so that's going to be 21. Okay, got it.

MS. IBRAGIMOVA: So the results are 67 percent high, 33 percent moderate, zero percent low, zero percent insufficient evidence.

CO-CHAIR SEPTIMUS: Okay Charlotte, if you'll take us through the gap.

DR. ALEXANDER: So for priority, only 28 percent of ICUs have a written policy, only 28 percent of physicians use maximal sterile barrier technique. Catheter-related bloodstream infections increase length of stay an average of 20 days with a cost of $3,000 to $60,000 per case. This involves a large number of line insertions with a significant morbidity and a high resource use.

CO-CHAIR SEPTIMUS: Question?

MS. ARDIZZONE: I just--I had a question when you said--you said 28 percent of the ICUs, you know that was a survey from 2002,

and some of the data is a little dated that was submitted. Do the developers have anything newer? Because I would think maybe in 15 years or so, practice has changed around maximal sterile barriers.

DR. DUTTON: We're certain it has, and in fact we're certain that's the reason why the CLABSI rate has dropped dramatically in the last few years, but there is no more recent published data on this. More anecdotally looking at our anesthesia practices that we went with and participate in the registry, there's been a substantial change in practice in this area in the last five to 10 years.

CO-CHAIR SEPTIMUS: Missy?

MS. DANFORTH: Yes, in terms of the performance gap, I'm just trying to understand, because we have--and we're going to be talking about it today--an outcome measure related to this, how you really judge--what's the significance I guess of the performance gap and the process measure if all the focus nationally

has been on the outcome measure?

DR. DUTTON: The importance of having both in this case is that the outcome measure is obviously what's most meaningful to patients and facilities and to the team as a whole, but in addressing gaps in the outcome, and obviously central line infections still happen, and they're still dangerous and expensive. In addressing a gap like that, as you heard Dr. Septimus say, there are multiple causes for central line infections, one of the more significant ones being care with how the line is placed, and that falls on a particular group of providers who are putting the lines in but may not be the ones managing them or using them for the long term. So we think at least that this is a situation where it's very appropriate to measure both the outcome and this component of the process, because it can give you real evidence for where to make improvements.

MS. DANFORTH: Is the measure then most appropriate for internal use for hospitals

and quality improvement, then?

DR. DUTTON: Well, this measure is obviously presented as a measure for providers, and we use it at multiple levels. Within the registry, it's reported at the level of anesthesia practices, at the level of facilities, and at the level of individual providers.

DR. BURSTIN: Just a brief comment. So--oh, I have so many microphones. So in general, you know, most of the measures we talk about, for example like the CLABSI rate, are at the hospital level. This is a clinician-level measure, so that's one of the disconnects. And so you still need clinician-level measures and we only--you know, a process measure in our parlance would really only be appropriate if there's a clear connection to outcomes. And this is one, you just discussed it in terms of the evidence, there's a clear process outcome link, and it's at a different level of analysis. So that's at least the logic of potentially maintaining some process measures that show directionality of how

to improve, even if you have the outcome measure at a different level.

CO-CHAIR SEPTIMUS: Okay, Rich?

DR. BRILLI: Rich Brilli. I don't mean to belabor it, but the developer continues to use the word CLABSI, and the measure talks about catheter-related, and as I said yesterday, they are not the same, and I think there needs to be some clear change in the language. Is this catheter-related? The vast majority of the adult hospitals, I was talking to Susan, actually report CLABSI, central line associated, which is the surveillance definition, not catheter-related, which is--I went through it yesterday so I won't repeat it. And in here it says catheter-related; he's saying CLABSI; I'm not sure which it is, so that's the first point.

Second point is I don't see an age issue here, and all of this applies to children, except premature infants where CHG may have significant skin problems, and they may use iodine alone and not chlorhexidine. So I think

that if it doesn't indeed apply to children, and I think most NICUs and ICUs, most children's hospitals are doing this already and have been doing it for a long time, with the exception of neonatal intensive care premature infants where CHG is not appropriate or may not be appropriate. So those are two things that need to be addressed somewhere in here.

DR. ALEXANDER: They did add the iodine as an appropriate skin prep.

DR. BRILLI: He said iodine with something else--

DR. ALEXANDER: Alcohol.

DR. BRILLI: Yes, so that still is not appropriate for a 1200 gram or 500 gram baby who has a PICC line put in; it needs to be just iodine alone. Their skin is very fragile and anything with alcohol or chlorhexidine may be problematic. So that's--and I didn't see an age limitation in here; it just said everybody. So I support it except for these premature infants.

CO-CHAIR SEPTIMUS: Okay, so the

question that I'm a little bit confused about-- oh, do you have a question? But it's just the fact is there's another bullet point here that says the reported performance scores show limited room for improvement, that the providers with non-zero reporting rates, Medicare patients aggregate by practice 2014 shows a mean performance of 94 percent. So where is that data from? And facility aggregate mean of 92 percent.

DR. DUTTON: Yes, so we have several different sources of data that we've examined for rates here. The Medicare five percent files are one; the National Anesthesia Clinical Outcomes Registry reporting rates are another. As I said, they're not exactly the same but running, as you point out, in the low 90s. We do have groups and there are facilities where the rate is close to 100 percent; we have other groups that are lower than that, but honestly from our perspective, the biggest gap in this is the many practices that still don't report it, and up to 40 percent of the central line placements we see in

administrative code sets like the five percent files, those measures are not reported.

CO-CHAIR SEPTIMUS: Okay, so we only have five more minutes, so I don't want to prolong, because I want Jason to give his comment, but the question based on the more recent data is, have we topped out in this, and has this becomes hard wired as the standard practice? I think that's the question that I think we have to ask ourselves when we vote on the gap. Jason?

DR. ADELMAN: Jason Adelman. I'm going to make some statements that I'm not sure are 100 true, so the developer can correct me, but the way I see this measure is that it's a great idea to--I support the idea of capturing the process measure that we follow proper technique; however, the way the developer describes where we are right now is that they've requested new CPT codes made that will capture all sorts of elements. Did you wear a cap? Did you wear a mask? Did you wear a sterile gown?

Did you wash your hands? This CPT code doesn't exist, and they say it will be approved --- it's supposed to be approved in August. And then so if it's--so it's a measure looking at capturing a CPT code that doesn't currently exist that we're voting on now, and then --- so you can't really test it, because it doesn't exist. And I would think that even if it does happen, it's not going to, in my opinion, do anything because you know, coders can only code if the doctors write all these things in the chart. And the only way a doctor is going to write, "I wore a cap and a mask and a sterile gown and sterile gloves and washed my hands" is if it's pre-written on some form, and when they sign that they've done it, they're just going to sign off on it, otherwise the coders--you know, it's not practically--and so I may be right, I may be wrong, and we can't even test it because the code doesn't even exist. So we're voting on a measure for something that may or may not work and I'm not sure if that's right or not, but I think that's how I read

what's going on, so I'll ask the developers.

DR. DUTTON: Actually, all of the codes--

CO-CHAIR SEPTIMUS: We'll ask the developer, and then we're going to vote.

DR. DUTTON: All of the codes have existed for four or five years; they -- we're changing some of them in small ways as I mentioned, to incorporate sterile ultrasound and to allow for the use of different prep solutions, include iodine, plain iodine. So the codes have been around; most of this data is captured from clinical documentation in the medical record, and when I put a central line in, I do record the use of maximal sterile barriers, appropriate prepping and draping.

DR. ADELMAN: Do you--sorry, do you write each element down or does your form sort of have--I mean there's a lot of elements that specify 2A1.

DR. DUTTON: Sure, and the question you're asking is obviously a moving target. At

the University of Chicago, I'm documenting this in Epic with a series of checkoff that might be completed by either me or by the operating room nurse who's observing the procedure.

CO-CHAIR SEPTIMUS: Okay we--

DR. DUTTON: But--

CO-CHAIR SEPTIMUS: --we need to vote because the timekeeper tells me we're out of time here already. So let's vote on--Lisa, go ahead. Lisa, go ahead, just do it real quick.

MS. MCGIFFERT: Okay, I'm looking at the skin prep directions here, and they seem to be saying--it says to use the chlorhexidine prep first, and if there's a contraindication to chlorhexidine, you go to tincture or an iodophor or 70 percent alcohol, and then it clearly says there's no comparison has been made between using chlorhexidine preps and the other preps, so it's an unresolved issue, and that no recommendation can be made for the safety or efficacy of chlorhexidine in infants, to address Rich's concern. So it's hard for me to know, to figure

out; it looks like still the first line is to use chlorhexidine or I mean the very first bullet says you can use any of these, and then the second and third bullet--the second bullet said you should use one first, and then if that's not indicated, you should use something else. So I'm just trying to get some clarity on the skin prep issue, and I know this is under evidence, and I'm sorry I didn't get that in earlier.

CO-CHAIR SEPTIMUS: Yes, we can talk about evidence. It's clear that chlorhexidine and alcohol is the preferred prep scientifically, so--but we'll get to that, but let's vote on the gap, okay? Do we think there's a gap in performance that would merit moving to the other discussions, okay? So, Laura?

MS. IBRAGIMOVA: Importance to measure and report 1B performance gap, the votes are 1 high, 2 moderate, 3 low, 4 insufficient.

CO-CHAIR SEPTIMUS: We got it.

MS. IBRAGIMOVA: So the results are 29 percent high, 38 percent moderate, 33 percent

low, zero percent insufficient.

CO-CHAIR SEPTIMUS: Okay then, we move forward. Charlotte?

DR. ALEXANDER: Specifications. The numerator is the patient for whom the central venous catheter was inserted with maximal sterile barrier technique that is CPT code 6030F; 6030F-1P as a documentation of medical reason for not using the maximal sterile barrier technique, including an increased risk of harm to the patient; 6030F-8P is the all elements were not followed. The denominator is all patients who undergo central venous catheter; the exclusion is 6030F-1P, which is where there was a contraindication. They have a proposed change to 6030F to add the ultrasound; it does not currently add ultrasound to it. Do you want to go into reliability yet?

CO-CHAIR SEPTIMUS: Yes.

DR. ALEXANDER: Reliability, this is captured through administrative claims, the Medicare Limited Data Set Carrier SAF five

percent and NACOR, which is the National Anesthesia Clinical Outcomes Registry. The performance level for NACOR, we had kappa of 0.97, for the SAF, the five percent Medicare was 0.95. I have a question, and I want to know how they're capturing the CPT-2 codes on non-Medicare, non-registry patients.

DR. DUTTON: We did capture--if they're not participating in either Medicare or the registry, we don't have any data from them, but that would be true of any measure. NACOR currently includes about 25 percent of all of the anesthesia practices and cases in the United States though, so it's a very large sample.

DR. ALEXANDER: So if there's a surgeon who's putting a line in a non-Medicare patient, how does he report it?

DR. DUTTON: The surgeon, sorry, the measure is specified for use by any physician who places a central line. It depends on the purpose that the measurement is being put to how the person would report it, whether at the local

level, the facility level, or to a national project. But surgeons could use this measure for PQRS, and this is a PQRS measure as well. If they report the appropriate codes, they would have to work out with their billing company or hospital system how to do the reporting. Many anesthesiologists do report this measure in PQRS and that's why the data is in CMS; many also report it to the registry; there's not a perfect overlap between those sets.

CO-CHAIR SEPTIMUS: Okay. Let us vote on reliability, please.

MS. IBRAGIMOVA: Scientific acceptability of measure properties 2A reliability, the votes are 1 high, 2 moderate, 3 low, 4 insufficient. The results are 14 percent high, 62 percent moderate, 24 percent low, zero percent insufficient.

CO-CHAIR SEPTIMUS: Okay, so let's go on to validity.

DR. ALEXANDER: For validity, this was done at the face value of a group of experts; the

score was 4.16 out of a 5 score.

CO-CHAIR SEPTIMUS: I'd say let's go ahead and vote.

DR. ALEXANDER: A meaningful difference--

CO-CHAIR SEPTIMUS: I'm sorry.

DR. ALEXANDER: --I'm sorry. Meaningful difference. There was a high performance rate among people reporting, but I had 60 percent; he stated at least 40 percent were not reporting, so of the people that report, they do a pretty good job, but there are a lot of people that are not reporting.

CO-CHAIR SEPTIMUS: Okay, let's vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties 2B validity, the results are 1 high, 2 moderate, 3 low, 4 insufficient. The results are 14 percent high, 67 percent moderate, 19 percent low, zero percent insufficient.

CO-CHAIR SEPTIMUS: Feasibility?

DR. ALEXANDER: This is captured

through administrative claims and electronic data through a clinical registry and uses CPT codes to capture the data.

CO-CHAIR SEPTIMUS: Seeing no hands, let's vote.

MS. IBRAGIMOVA: Feasibility, the votes are 1 high, 2 moderate, 3 low, 4 insufficient. The results are 33 percent high, 38 percent moderate, 29 percent low, zero percent insufficient.

CO-CHAIR SEPTIMUS: Okay, then we go to usability.

DR. ALEXANDER: This is currently being used for PQRS, for the anesthesia registry, and there's some discussion about Joint Commission using it also in evaluation of hospitals.

CO-CHAIR SEPTIMUS: Discussion? Yes.

MS. ARDIZZONE: Just one comment. I agree this is really important because it identifies something--because there's two components of the line: the insertion, and the

maintenance, and this really gets at the insertion. The only thing I'm concerned about is in many places, this is just self-reported, that I've done maximal sterile barrier precautions. It might be better if, you know, someone was actually observing that someone was doing maximal sterile barrier precautions. I don't know how feasible that is to have somebody observing, an independent observer watching everybody. So I'm concerned that it's just self-reported, but--

DR. DUTTON: Yes, this is Dr. Dutton again. Us too. We get that. First, in many institutions, and depending on the documentation system, it actually is an observer documenting this. I mentioned the OR nurse, it can also be the ICU nurse documenting this for the procedure. You can argue about how independent they are; it's still a checkbox exercise. And this is why we're working with this measure, as with all of our measures, on e-specifications, and hope to come forward next year with an eMeasure on the same topic. As you heard earlier, it won't

exactly match this one, but we believe it will be able to get the same data or similar data in a much more objective way.

CO-CHAIR SEPTIMUS: Okay. Usability. Let's vote.

MS. IBRAGIMOVA: Usability and use. The votes are one high, two moderate, three low, four insufficient information.

CO-CHAIR SEPTIMUS: One more.

MS. THEBERGE: Ann, I need your vote.

CO-CHAIR SEPTIMUS: All right, we've got 21.

DR. APPLEGATE: Ed, this is Kimberly, could I make a comment about that last comment? A concern about the observation?

CO-CHAIR SEPTIMUS: Of course, Kimberly. Wait one minute, because we're trying to get the one last vote and then, you can comment.

DR. APPLEGATE: Well, I'd like to make it before the vote.

CO-CHAIR SEPTIMUS: Well, the vote --

DR. APPLEGATE: I raised my hand.

CO-CHAIR SEPTIMUS: The vote's been cast, Kimberly, I'm sorry.

DR. APPLEGATE: That's all right.

CO-CHAIR SEPTIMUS: And I know it's difficult for you on the phone, so if you would text Suzanne, and we'll make sure that you -- oh, so you missed it? Okay. Well, I'll tell you what. Since we have to vote again, why don't you go ahead then and tell us your comment.

DR. APPLEGATE: I just wanted to second the author's comment that when the healthcare -- it's an awful lot of line placements and oversight and review of infections CLABSI by interventional radiology, he puts in an enormous number of these central lines and PICCs honestly, when the nursing teams can't put them in. And so, I review a lot of these cases and try to ascertain if it's secondary to placement or dwell. And so, the healthcare system tries very hard to say everyone does it the same, and it's just, at least at Emory, a bundle.

So, you know, yes, we'd love to have observations, and we do do a little bit of observation. But basically, it is a check right now, and we try to audit some of the checks, internally, and in our department, a very large department, but I would say that just by having the bundle and having it across all systems with everybody doing it the same and everyone given the bundle, I think that that's the way to address this concern that was raised by someone. I don't know who it was.

CO-CHAIR SEPTIMUS: Thank you, Kimberly. All right, now, let's vote again on usability. That was fast.

MS. IBRAGIMOVA: So, the results are 23 percent high, 59 percent moderate, 18 percent low, zero percent insufficient information.

CO-CHAIR SEPTIMUS: Okay. We're going to go to the last one. I'm going to just make a comment, not as co-chair but as someone who knows a little bit about this area. We've focused a lot of our attention on insertion, but I want to

tell you that most of the more recent studies in terms of prevention emphasize maintenance. This has nothing to do with maintenance and again I'll just come back to my original comment: I think we're seeing a higher level of compliance. I think compliance should be, by the way, should occur. I think maximum barrier in putting this in carefully is very, very important, but I just want to put this in relative --- as to what the most salient points are in 2015 to prevent CLABSIs, just to put that in perspective.

Okay. Let's vote on whether or not it's suitable for endorsement.

MS. IBRAGIMOVA: Overall suitability for endorsement: does the measure meet NQF criteria for endorsement? One yes, two no.

CO-CHAIR SEPTIMUS: Sorry, did I miss you again, Richard?

DR. BRILLI: Well, I don't think I had my card up in time. I think there's something important on page ten, there's a competing measure which talks about CLABSI as opposed to --

so I'm not sure -- page ten of the document that's submitted, it talks about competing measures. And there's a whole other measure that this group has already approved, which talks about CLABSI, C-L-A-B-S-I. And this is -- I don't know if it's the exact same measure, this is just C-R-B-S-I. So again, it's the same -- I'm not quite sure what we're doing here that sounds like it's completely --

CO-CHAIR SEPTIMUS: Well. Well, I think it actually was explained by Helen very well that this is a process measure, other one's an outcome measure.

DR. BRILLI: Is there no process measure in the CLABSI measure?

CO-CHAIR SEPTIMUS: No, there's no process measure, and this is at the clinician level and not at the hospital level. So they're different measures.

You want to vote again? I think we need to vote again. Sorry, let's vote again.

MS. IBRAGIMOVA: Yes, we need four

more votes. Just one more.

CO-CHAIR SEPTIMUS: Can we tell, Laura, who hasn't voted? Did Missy vote? Okay.

MS. IBRAGIMOVA: So results are 86 percent yes, 14 percent no.

CO-CHAIR SEPTIMUS: Well, thank you for the developers and thank you for an excellent discussion, Charlotte. And now we're going to move to a new measure. Are the CDC developers on the phone?

MS. ARDIZZONE: Are we going to talk about retirement, or?

CO-CHAIR SEPTIMUS: Retirement? Reserve status?

MS. ARDIZZONE: Yes, that. Sorry.

DR. POLLOCK: Yes, Ed, this is Dan Pollock, I'm on the phone.

CO-CHAIR SEPTIMUS: Yes, one second, Dan. So --- since it passed gap, it doesn't go into reserve. Okay. All right, so the next measure -- we may have to cut our break short here, but that's okay -- is a new measure from

NHSN on antimicrobial use measure from the CDC. And I think Dan Pollock is going to be discussing it as a developer; is that correct, Dan?

DR. POLLOCK: That's correct.

CO-CHAIR SEPTIMUS: Okay. The floor is yours.

DR. POLLOCK: Okay. Thank you so much. I'm Dan Pollock. I'm a medical epidemiologist at CDC working with colleagues on the National Healthcare Safety Network, N-H-S-N. We have a proposed measure of antimicrobial use. Antibiotic overuse or inappropriate use in U.S. hospitals is a widely recognized clinical and public health problem that places individual patients at risk for adverse outcomes, increases the incidence and prevalence of antimicrobial resistance, and jeopardizes the effectiveness of a vitally-important healthcare resource for the general population.

CDC estimates that at least two million people become infected with bacteria that are resistant to antibiotics, leading to 23,000

deaths, and recommends strongly improvement of antibiotic prescribing as a core action to prevent resistance.

Numerous individual studies and systematic reviews provide strong evidence that measurement of antimicrobial use and data-driven interventions by antimicrobial stewardship programs, or ASPs, lead to more judicious use of antibiotics, reduced antimicrobial resistance, and other favorable healthcare outcomes.

So the NHSN AU measure proposal is one that really seeks to provide data, benchmarks of antimicrobial use at the national level for stewardship programs to use in their systematic efforts to guide prescribing practices. The AU measure provides summary results that hospital and health system ASPs can use as quantitative aids. The core metric is the standardized antimicrobial administration ratio, or SAAR, and we focus in that measure on the ratio of observed to predicted antimicrobial use.

The SAAR is focused on high-value

targets for stewardship programs and high-level indicators of antibiotic use for ASPs. The SAARs can be used by ASPs to benchmark antimicrobial use in multiple patient care locations, identify opportunities for improvement, and gauge impact of stewardship efforts.

At the outset, the SAARs provide a set of signals that often warrant further analysis, such as an evaluation of the extent to which a specific antibiotic or group of antibiotics accounts for a high SAAR value and the extent to which an antibiotic or group of antibiotics were used appropriately. The SAAR, in and of itself, is not a definitive measure of appropriateness. That requires additional information.

Some of the analytic follow-up can be completed with hospital and patient care location-specific data reported to CDC's National Healthcare Safety Network using analytic features built into the application. However, additional analyses to determine the appropriateness of antibiotic use in individual instances are likely

to require access to detailed patient-level data that is beyond the scope of data collection and analysis using the NHSN antimicrobial use and resistance module, such as clinical indications for specific antibiotics and dose duration decisions. Those are not part of our surveillance.

The measure relies completely on electronic data using for the numerator, medication administration data that are ascertained via electronic medication administration record systems or barcode medication administration record systems with the denominator data coming from ADP systems. We've worked closely with five vendors and a homegrown system, as well as stewardship programs and three leading healthcare systems nationwide to implement, develop, and preliminarily use data from the system.

We're proposing a measure for public health surveillance for quality measurement and improvement. We are decidedly not proposing this

measure for public reporting or payment purposes until we gain greater experience with the measure and provide additional information around the predictive modeling, grow that predictive model in ways that would allow appropriate use for public reporting and payment purposes.

With that, I'll stop.

CO-CHAIR SEPTIMUS: Thank you very much, Dan. Charlotte, are you on to discuss this one again? Boy, Charlotte, so --- Charlotte's going to walk us through the data elements, Dan, and then we'll take questions from all. And please stay on the line because I'm sure we'll have some questions for you. Charlotte?

DR. POLLOCK: Sure.

DR. ALEXANDER: So, I think he gave a pretty good description of data elements. I'm going to go to the evidence. Clinical practice --- this uses a clinical practice guideline, which recommends prospective audit and feedback, formulate a restriction, computer-based surveillance to target antimicrobial

interventions, resistance patterns in nosocomial infections, and adverse drug events. Up to 50 percent of antibiotic use is inappropriate, leading to the development of increased drug resistance, such as carbapenem-resistant Enterobacteriaceae and C. diff.

Several systematic reviews demonstrated that stewardship programs can lower the risk of C. diff and other outcomes. They did not specifically test this measure, but the quality, quantity, and consistency is high.

CO-CHAIR SEPTIMUS: Lisa?

MS. MCGIFFERT: Yes. Hi, Dan, it's Lisa McGiffert. I had some questions about the measure, not about evidence, so much, but the measure.

CO-CHAIR SEPTIMUS: I'm sorry.

MS. MCGIFFERT: Is this the time to talk about it?

CO-CHAIR SEPTIMUS: No, we're only talking about evidence.

MS. MCGIFFERT: Well, when will we

talk about the measure?

CO-CHAIR SEPTIMUS: As we go along.

MS. MCGIFFERT: Okay. What I would like to know is what the SAAR, can you talk to me about the SAAR a little bit more? I see that it's a one is the score that people would be going for, and I'm trying to figure out is one an average? How did you determine that is the optimal score? Can you tell us a little bit more about that?

CO-CHAIR SEPTIMUS: Lisa, that is going to be -- that's a reliability question. So, if you can hold that thought, let's go through the evidence and then, we'll get to the relia-- that's an excellent question, but it's like a SAAR, but it's for antibiotics. But let's wait until we get to the reliability. Let's go through the evidence first. I'm sorry to hold off on that question.

Anything else? Yes, Pat?

DR. QUIGLEY: Thank you, Dr. Septimus. This is Pat Quigley presenting, Dr. Pollock, and

I hope it would be acceptable in relationship to the evidence to share that I have recently had recent conversations with Dr. Ann Hendrich, nurse extraordinaire of Ascension Health, and she had shared that there was a recent meeting at the White House with the three major healthcare systems in launching this leadership stewardship with CDC. And while there may be evidence related to the work that they're doing already, and these three healthcare systems are not in the public domain in terms of the literature.

So, maybe this is part of public reporting, that maybe you could do a little summary of that White House Conference Aging to help inform this body? Or the White House Conference on -- I was at the White House Conference on Aging, I'm sorry. But the White House conference on this measure.

CO-CHAIR SEPTIMUS: You mean the White House Summit on Antimicrobial Resistance that was just held?

DR. QUIGLEY: Yes. Yes. Just held

like two weeks ago I think.

CO-CHAIR SEPTIMUS: I was at the conference!

DR. QUIGLEY: You were there, huh?

CO-CHAIR SEPTIMUS: But I'm just --- you said about the measure specifically?

DR. QUIGLEY: It was a White House meeting. It was a meeting at the White House.

CO-CHAIR SEPTIMUS: About this specific measure?

DR. QUIGLEY: Yes. Yes.

CO-CHAIR SEPTIMUS: I'll let Dan answer that.

DR. POLLOCK: Well, actually, Ed, I'm going to toss it back to you, because I was not at the Antimicrobial Stewardship Forum that was at the White House. But I understand that it was a very successful meeting with many national organizations stepping up, articulating their commitments to stewardship efforts to reporting into the Antimicrobial Use and Resistance Module. And we're expecting a lot of positive energy to

come out of that process.

DR. QUIGLEY: Thank you.

CO-CHAIR SEPTIMUS: They didn't have a detailed discussion about this specific measure. But certainly, in the White House report that was published in March, certainly getting facilities to report into the AU Module, was it by 2017, Dan?

DR. POLLOCK: Correct.

CO-CHAIR SEPTIMUS: It's certainly a high priority to increase the number of facilities that are already doing --- just like they are for HAIs. But in terms of specific measures, it's assumed that measures will come forth through NQF. But actually, this particular measure was not discussed at the White House summit.

DR QUIGLEY: Okay, I apologize.

CO-CHAIR SEPTIMUS: That's fine.

DR. BURSTIN: Although, I will say that we've been actually outreaching to CDC and others. And given this huge interest in

antimicrobial stewardship, and how important it is, we actually have -- we were delighted CDC was willing to bring forward some of their newer measures in this area, where we, frankly, have none.

MS. MCGIFFERT: This is actually included in the Administration's National Action Plan.

CO-CHAIR SEPTIMUS: Well, measures like this. Yes, that's correct.

DR. YU: Yanling Yu. I have a question about denominator and nominator. Is this include -- should it include -- under reliability thing?

CO-CHAIR SEPTIMUS: It's reliability, also.

DR. YU: Thank you.

CO-CHAIR SEPTIMUS: Steve?

DR. LAWLESS: Yes, it is just a question of reliability, or not. In the numerator piece, is this antibiotic days captured with a positive culture or, just --

CO-CHAIR SEPTIMUS: Hold that question, okay? All right, so, Yanling, did you have another question? No. Okay. So let's vote.

MS. IBRAGIMOVA: Importance to measure and report 1(a) evidence, structure, process, intermediate outcome. The votes are one high, only eligible if QQC submitted; two moderate; three low; four insufficient evidence.

And the results are 68 percent high, 23 percent moderate, 5 percent low, 5 percent insufficient evidence.

CO-CHAIR SEPTIMUS: Gap. Charlotte?

DR. ALEXANDER: I'm going to make one little statement to help clarify Lisa's question and just as I go into gap. This measure is antibiotic use reported to the CDC for adult and pediatric patients compared to predicted on the basis of nationally-aggregated data. SAARs summarizes the observed to predicted antimicrobial use for one of 16 antibacterial patient-care location combinations.

If the measurement equals the SAARs, it's a one. If it does better than SAARs, it's negative. If it does more than SAARs, it's positive.

DR. POLLOCK: This is Dan. I just need to correct that. There are no negatives here. The value of one would mean that the observed to predicted is equivalent. A value higher than one would mean that the observed to predicted is higher, the observed antimicrobial use is higher than would be predicted. And, similarly, if the observed is below one, it would be indicative of antimicrobial use that was less than predicted.

CO-CHAIR SEPTIMUS: So does everybody know what a SAAR is before we -- just to make sure everybody -- okay. And just to let you know, in HAIs, it's a standard infection ratio. Again, it's observed over expected based on certain risk adjustments. But, Lisa, you had a question about SAAR?

MS. MCGIFFERT: Okay. Dan, me again.

Lisa. So the SIR, the predicted, we know where the predicted came from. Can you tell us where the predicted comes from in the SAAR, please?

DR. POLLOCK: Sure. So the predicted is the days of therapy that would be predicted from the data that are available to CDC, that are aggregated by CDC, that are used in conjunction with a predictive model to produce a summary of what would be predicted for patient location or facility-wide. So we have a statistical process that's applied to the actual data that are reported in.

To construct the SAARs, we use data from over a hundred healthcare facilities that were reporting from 25 states. And this is an initial group of early adopters whose data were used to develop the predictive models that comprise the denominators for each of the 16 SAAR metrics, which, as I said earlier, are a combination of antimicrobial categories with patient-care locations.

So an example of a particular SAAR

would be anti-MRSA agents used in adult ICU locations: medical, medical/surgical, and surgical. That would be a single one of the 16 SAARs.

MS. MCGIFFERT: And that would be based on what you would ideally want to see in the hospitals. So you took the 100 hospitals and you looked at appropriate use.

DR. POLLOCK: Let me clarify, Lisa. This is not a statement of ideal. This is not a statement of appropriateness. This is a statement of what would be predicted, what's going on nationally for the particular patient care location.

MS. MCGIFFERT: So let me ask you this: what is going on nationally, is that a good thing or a bad thing?

DR. POLLOCK: We need more, we need more data with which to address that. And, certainly, that's part of what CDC and partner organizations are pursuing is to, you know, ultimately drive down the overuse of

antimicrobial agents, which leads to antimicrobial resistance and other bad outcomes. So this is an integral part of that effort, but it is not the singular solution to driving down overuse and improving the prescribing practices. However, it does provide a mechanism for the first time to enable national benchmarks to be available to literally thousands of hospitals throughout the country that enroll and participate in the module.

MS. MCGIFFERT: Okay. Well, I'll just say one more thing because, you know, we've been around this road before, about using the terms "benchmarks" because when people see benchmarks they see targets. That's what they want to strive to achieve. And I guess a benchmark is a little bit better than using the word "target." But, you know, just to be clear, it's just going to be an observed where everyone is, rather than a goal.

And I know you're not suggesting this for public reporting, but I sure am going to

suggest you use it for public reporting. And I think it's, you know, we're just going to have to be real clear about what, you know, what the measure shows us. So thanks.

DR. POLLOCK: Again, I agree with you, Lisa. But I think we're, we have to be clear about it. And I think, eventually, you know, we'll be very supportive of a next iteration of this measure for public reporting and payment purposes. But it's a new measure. It reflects what we had worked very closely with stewardship programs on over a period of years, and we're confident that this is a very important first step to take. But it is just a first step.

CO-CHAIR SEPTIMUS: Richard?

DR. BRILLI: Hi, Rich Brilli. Just a question for the developer. Since this applies to kids, is there sufficient data, predictive data for children that this should apply there, not that this is going to be used as a benchmark? We certainly don't use it at our place, and I've got a number of other pediatric people I've

communicated with that don't use it. So do you have sufficient pediatric predictive data to use this here?

DR. POLLOCK: We believe that we do. And I think one of the reasons why it's not used, yes, it's because it is a new metric, it is a new measure. But we've worked closely with stewardship programs, including pediatric stewardship programs, to develop the measure. We've got a separate set of SAARs entirely for pediatric patient care locations. We are confident that what we are developing will be relevant for both the adult and the pediatric patient populations.

Neonates are another story. That one is on the horizon, but we have a strategy for bringing the neonatal population in, as well.

DR. BRILLI: Okay, thanks.

CO-CHAIR SEPTIMUS: Okay. I think it's time to vote. Laura?

MS. IBRAGIMOVA: So importance to measure and report 1B, performance gap, the votes

are one high, two moderate, three low, four insufficient. Just two more votes.

DR. RISING: Please don't count me in for now. Since I just arrived, I'm going to sit this vote out. Thank you.

CO-CHAIR SEPTIMUS: We would never discount you, Josh.

MS. IBRAGIMOVA: So the results are 59 percent high, 32 percent moderate, zero percent low, 9 percent insufficient.

CO-CHAIR SEPTIMUS: Okay. Now we get to the some of the questions that some of you were raising with reliability. So Charlotte?

DR. ALEXANDER: Under scientific acceptability, the numerator is the days of antimicrobial therapy for antimicrobial agents administered to adult and pediatric patients in medical, med/surg, and surgical wards and medical, med/surg, and surgical ICUs.

Specific measurements are, one, broad spectrum antibiotics for hospital onset multi-drug-resistant infections; two, broad spectrum

antibiotics for community-acquired infections; three, anti-MRSA agents; and, four, surgical site prophylaxis agents; and, five, all antibiotics.

The denominator is days present for each patient care location defined as any portion of a day in a calendar month. All days are summed for each location and month, and the aggregate sums comprise the denominator.

Exclusions are locations other than those stated above. The data is stratified by hospital and patient location-specific variables, teaching status, hospital bed size, ICU status, ICU bed size, patient care location and bed size. It is risk adjusted using a negative binomial regression model to find factors associated with differences in use rates and to predict days of therapy that can be compared to observed days. SAAR is the ratio of observed to predicted antimicrobial use less than, equal, or greater than one. I have a concern that patient days may be double-counted if you have transfers since they're being counted any time a patient is in a

location.

Reliability testing. This is tested at the level of the facility. There were 24 hospitals used for the aggregate and 13 for the data. These were hospitals who were reporting to CDC NHSN Antimicrobial Use and Resistance Module over the years 2011 to 2014.

On the data elements, there was a 60 to 80-percent reliability and greater than 99 percent on the process. And my only comment is it is a small sample for data.

CO-CHAIR SEPTIMUS: Do you want to, you raised a question about double counting.

DR. ALEXANDER: Yes. I think --

CO-CHAIR SEPTIMUS: Dan, do you want to, how you count days? Is there a --

DR. POLLOCK: Right. So, yes, a portion of a day in a patient care location where an antimicrobial agent is administered counts as an antimicrobial day for that patient care location. So if a patient is transferred from a ward location to an ICU location, for example, or

vice versa, if they're maintaining that antimicrobial agent throughout, that antimicrobial agent administered for a portion of the day in each of those locations would be counted as an antimicrobial day in each of those locations.

CO-CHAIR SEPTIMUS: Okay. Questions? Steve?

DR. LAWLESS: Yes, this is Steve Lawless. A couple of questions. One is I've looked at the references that you provided, so thank you for that. But there's nothing in the references that would actually show anything about the data. All the references are mostly about association with C. diff, and I get that. But in terms of the general distribution of SAARs, where they are, the details and the numbers and the risk adjustments. We're used to seeing a lot more of that data out there, percentile rankings. Where is that?

And the other piece is --

DR. POLLOCK: Well --

DR. LAWLESS: Okay, go ahead.

DR. POLLOCK: Well, good question. I mean, again, this is a new measure. We've provided in table three that accompanies the measure proposal the SAAR distribution and statistical comparisons for each of the 16 SAAR metrics. And, admittedly, this is the first time that the SAAR is being reported.

But, again, it's grounded on concepts that have existed in the stewardship domain for many years, albeit with specifications that we've had to tailor to fit with NHSN and what NHSN can produce so that, yes, it is novel. But you have to start somewhere.

DR. LAWLESS: Right. And the next question would be is just consider the idea of positive cultures versus non-positive cultures as a comparative so that overuse of antibiotics a lot of times is, you know, we have this classic 7, 14, 21 days of antibiotic courses grounded in no particular fact. But if something as positive as a culture may be appropriate versus no

cultures for just overuse, are we considering that as you're developing that?

And the third piece is what about the outpatient? A lot of patients do come into the hospital already started on an antibiotic, and so people tend to continue the antibiotic because you just don't know what to do or not, whether it's working. Is that a consideration or not?

DR. POLLOCK: So very good observations and comments. I would say that the presence or absence of a positive culture certainly would be part of a consideration that a stewardship program would ultimately incorporate in evaluating a SAAR value at the patient level. But we're not ascertaining the presence or absence of a positive culture as part of the routine surveillance effort.

We do have a companion piece to the antimicrobial use reporting, which is the antimicrobial resistance part of our AUR module, and that will provide at least some indication of the amount of bacterial culturing that's going on

and the results that are being ascertained. But, again, that will be not at the patient level, that will be at the location and the facility level.

Your other question is a very good one. This is a starting place. The patient care locations that we have selected can be expanded out to include emergency department and other locations, but we wanted to begin with some targets that we think are important. Not to say that there aren't other targets, but we want to learn from these initial set of targets and then propose expansion of the coverage, both in terms of patient care locations and other facility types. These are hospital participants, and the measure is really focusing on hospital antimicrobial use.

CO-CHAIR SEPTIMUS: Okay. One last comment, and then we're going to vote. Yanling?

DR. YU: Yes, I have a couple of questions. One is there's a reliability test, the paper records were used to analyze the model

and the numerator. But in the measure, it says only electronic records would be used. I'm just wondering if you have any source of explanation why the paper records would not be considered eventually?

DR. POLLOCK: Well, that's a good question. I think that we, like others, want to move forward to electronic quality measurement and using electronic supply chains of information. We have in the past attempted to capture a true NHSN antimicrobial use with manual processes and manual data entry. It's simply proved to be untenable operationally. The good fortune that we have is that there is rapidly-increased use of the electronic bedside medication administration record-keeping systems. So really our operational design pivots off of use of those systems, extracting, transforming, and loading data that have been ascertained through those systems into a message that could be sent to CDC. Again, that's electronic supply chain of information.

The reliability testing that we are doing is reliability testing before the messaging begins, and it looks at the data that are put into the message compared with the data that are in native systems, be they the medication administration systems in the case of the numerator or the admission discharge transfer systems in the case of the denominator.

DR. YU: Okay. Thank you for your explanation. My second question is, in your regression model, you called standard population. I just wondered if you have any examples about this standard population? Are those theoretical ones, or do you just gather real data and then start to, you know, to characterize those different population?

DR. POLLOCK: It's the latter. I mean, we're using real data. We're using the nationally-aggregated data, and the initial adoption and use of our AU reporting that we use to develop a measure includes data from over 100 healthcare facilities nationwide from 28 states:

critical access hospitals, children's hospitals, an oncology hospital, in addition to the predominant general acute care hospitals that participated. We've used multiple vendor systems and a homegrown system.

So we have, I think, quite heterogenous participation in the AU reporting already. That's going to grow. It's growing considerably, even since we introduced a bit of a measure proposal in April. And as we further the participation in the module, we can use these additional data in the modeling process.

DR. YU: Okay. So help me understand. So SAARs really is a model --

CO-CHAIR SEPTIMUS: It's SAAR. It's SAAR, not SAARs, please.

DR. YU: Yes, SAAR. It's really a theoretical number; is that right?

DR. POLLOCK: Well, I'm not sure what you mean by --

DR. YU: It's optimal days compared with what's observed, the ratio.

DR. POLLOCK: Well, it doesn't have to do with optimal. It doesn't have to do with ideal. It has to do with what's going on nationally. It doesn't assume what's going on nationally is appropriate. That's why it requires additional analysis at the institution level to really get at questions of optimization of the antimicrobial prescribing.

Now, over time, when we had the opportunity to add to our model, perhaps using data from the antimicrobial resistance reporting, we'll be in a better position to get closer to an understanding of where, on the basis of the SAAR alone, antimicrobials are being overused.

DR. YU: Okay. All right, thank you.

CO-CHAIR SEPTIMUS: Okay. We're going to vote on this. We have to move fairly quickly here, but I want to allow enough time because this is a new measure. So let's vote on reliability.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2A,

reliability. The votes are one high, two moderate, three low, four insufficient.

CO-CHAIR SEPTIMUS: Josh, you're not voting, right? Okay.

MS. IBRAGIMOVA: The results are 26 percent high, 61 percent moderate, 4 percent low, and 9 percent insufficient.

CO-CHAIR SEPTIMUS: Validity. Any major comments on validity, Charlotte?

DR. ALEXANDER: I think it's important to point out that the evidence refers to the value of a stewardship program and audit feedback, not directly to reporting. Face validity was done with an expert panel, and this is risk adjusted with a statistical risk model.

CO-CHAIR SEPTIMUS: Yanling, is that still up, or are you finished with it? Okay.

DR. YU: I apologize.

CO-CHAIR SEPTIMUS: No, don't apologize. Okay. Let's vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2B,

validity. The votes are one high, two moderate, three low, four insufficient. Just need one more vote.

The results are 30 percent high, 57 percent moderate, 4 percent low, 9 percent insufficient.

CO-CHAIR SEPTIMUS: Okay. We'll move to feasibility.

DR. ALEXANDER: This measure is captured through electronic data during the provision of care.

CO-CHAIR SEPTIMUS: Yes, Leslie?

DR. SCHULTZ: Leslie Schultz. Dan, I have a question. There are probably more hospitals that are not fully e-enabled than there are hospitals that are. Is there a proxy for those who cannot easily get to the administration data?

DR. POLLOCK: At this point, we think that there is overwhelming movement to electronic medication administration systems or barcode systems. And we've got survey data that's

indicative of that. We have critical access hospitals that are using this type of technology.

So it's really ubiquitous. And rather than introducing a manual process which we've had experience in the past it leads to frustration and inability to report, we've placed a stake in the ground here in saying we want to go electronic. And we think this is a good place to begin. This is, indeed, while not meeting the criteria of what is defined as an eMeasure, it's still an electronic measure, and it's time for us to move forward and move from here.

CO-CHAIR SEPTIMUS: Lisa?

MS. MCGIFFERT: Dan, this is Lisa McGiffert. Did you say that you had a homegrown version of electronic reporting?

DR. POLLOCK: Right.

MS. MCGIFFERT: And so does that -- and just to follow up, does that mean that CDC, that's something CDC could make available to smaller hospitals that might not be able to afford a private contract?

DR. POLLOCK: Well, the homegrown system is a system that was developed and used to extract the data from an electronic system that is commercial. So it's, in essence, analogous to a third-party software system, which is what the vendors are providing. And, yes, we hear, you know, it's possible to have a vendor system deployed for purposes of extract, transform, and load into a message to send to CDC. But when you look at the dollar amounts that are being discussed, $25,000 to $50,000 for these types of implementations, and you compare that dollar amount with what IT budgets are in hospitals or what's been expended societally on the meaningful use program and what the return on investment, the $25,000 to $50,000 a year for stewardship efforts, we think that this is, again, a place to put a stake in the ground and say this should be part of the cost of doing business in American healthcare.

CO-CHAIR SEPTIMUS: Okay. Let's vote on this then, please. Feasibility.

MS. IBRAGIMOVA: Feasibility. The votes are one high, two moderate, three low, four insufficient. The results are 22 percent high, 65 percent moderate, 4 percent low, 9 percent insufficient.

CO-CHAIR SEPTIMUS: Okay. Now we get to usability. Charlotte?

DR. ALEXANDER: This is not currently a reported measure. It's a new measure, but it is intended to be reported in the National Healthcare Safety Network and used for internal and external quality improvement and benchmarking.

CO-CHAIR SEPTIMUS: Missy? Haven't heard from you in a while.

MS. DANFORTH: Hi, Dan. This is Missy Danforth from the Leapfrog Group. I just have a quick question. I know that data is going to be reported through the electronic reporting system. I'm wondering, because you've acknowledged this as a new measure, not an ideal measure, you're still learning a lot. Is there an opportunity to

incorporate the CDC's checklist that was developed for antimicrobial stewardship into the reporting on this measure so we can get a better sense of what are the elements on that checklist that are most tied to lower, more appropriate use, since the checklist is really expansive? It's something that the CDC has developed and recommending that all hospitals use. Is there a way to incorporate that into the reporting, as well, so that we can get as much information as possible and possibly improve on the measure faster?

DR. POLLOCK: That's a good question. I think that the way -- and we certainly have that in our sights, as well, but not so much to incorporate it into the measure but to use the measure results for antimicrobial use reporting, as well as the survey results of what elements of the stewardship program are present and reported by a facility through a NHSN annual survey. We can use those in conjunction with each other in analyzing the relationship, and that certainly is

part of our strategy going forward.

DR. BURSTIN: And just to add this on, we've also had conversations with Arjun at CDC about that. And I think there may be an opportunity as they get back the survey data this year from 4,000 hospitals or something like that to think about how to build that into a measure to follow.

DR. POLLOCK: That's right, Helen. That's exactly right.

CO-CHAIR SEPTIMUS: There's been some discussion about a process measure built around that. Also, the TATFAR report, Transatlantic Taskforce on Antimicrobial Resistance, will release its structure and process measures any moment, right, Dan?

DR. POLLOCK: Let's hope so.

CO-CHAIR SEPTIMUS: It's been approved, but that will also be out there. But we have discussed looking at process measures as another potential measure for stewardship. Leslie?

DR. SCHULTZ: Leslie Schultz. A comment here. It is a new measure. I think it's going to be a wonderfully-helpful measure. We need something standardized that's endorsed so we can grow that hundred to forty-five hundred and really understand what is going on. What are we using? We can't talk about overutilization until we see what that distribution is. And having done work with CDC on potentially inappropriate overutilization of intravenous antibiotics, I think this is very important. We as a nation need to measure so we can manage.

CO-CHAIR SEPTIMUS: Steve? One last question?

DR. LAWLESS: Yes, just a question of NQF processing protocol. I think everybody is realizing the importance of this, the value long-term. It's something that really has a lot of development to be done. I mean, the value is there. As we endorse this, it's lots of measures that need to be done, future success of everything is going to be independent. Our

endorsement will do what for this measure, versus they would develop it and come back later with a more mature, guys, we have something we proudly endorse versus we don't know where it's going.

DR. BURSTIN: I think that's a great question. This is Helen for Dan on the phone. I think our feeling is that we want to get something out in a space. It's really important, it's a national priority, and, frankly, we don't feel like there's enough there. I think putting it out in a space makes it clear this is important in the measurement enterprise. Hopefully, people begin using it. And, again, as you've seen, it's not as if when we endorse something it stops the development train. I think there's a lot of momentum here to keep making new measures, make them better. It also potentially makes it something that could be picked up as part of some of the federal infrastructure around payment or public reporting, not immediately but at least puts it on their radar screen perhaps, I think, in a way

that it may not be if it doesn't kind of go through this national process.

CO-CHAIR SEPTIMUS: That's a very good question, Steve. Okay. One more question and then we've got to -- go ahead, Lisa.

MS. MCGIFFERT: Okay. I just would say that the history with NHSN and CDC has certainly been that when these measures have first been put out, even when they first started collecting infection information, it has significantly changed over time through input from the users and looking at the data. And I expect we'll see a lot of that in the future, too.

CO-CHAIR SEPTIMUS: Okay. Let's vote on usability, please.

MS. IBRAGIMOVA: Usability and use. The votes are one high, two moderate, three low, four insufficient information. Just two more votes. The results are 39 percent high, 48 percent moderate, 4 percent low, 9 percent insufficient information.

CO-CHAIR SEPTIMUS: Okay. So the last vote then is suitability for endorsement. So it's either yes or no.

MS. IBRAGIMOVA: So overall suitability for endorsement, does the measure meet NQF criteria for endorsement? One yes, two no.

CO-CHAIR SEPTIMUS: One more.

MS. IBRAGIMOVA: So the results are 91 percent yes, 9 percent no.

CO-CHAIR SEPTIMUS: Thank you, Dan, for being on the phone. This is a historic first step, I think, to addressing antimicrobial resistance, and we certainly expect over time that we'll see other measures and refinement as more institutions report in to the AU Module. So thank you for all the effort it took in bringing this forward and thank the Committee for considering this. So thank you very much, Dan.

DR. POLLOCK: Thank you and thanks to the Committee.

DR. BURSTIN: Thanks, Dan.

CO-CHAIR SEPTIMUS: When I get the TATFAR report, I'm more than willing to share. I have it on my computer, but it has not been publicly released. I'm sorry. I can't do it.

Okay. The next on our agenda is 2729, timely evaluation of high-risk individuals in the emergency department. CMS and Mathematica -- I am going to recuse myself from voting on this one, so if you see 23 votes, Laura, it means I shouldn't have voted. But I'm not going to vote on this because I was one of the consultants of the development of the measure. Tom, are you going to do this? Cindy, okay. So if you'll introduce yourself to the Committee, and the discussant on this is, oh, Kendall. Thank you. Go for it.

MS. CULLEN: Great, thanks. Good morning. My name is Cindy Cullen. I'm the project director for the CMS Hospital Inpatient and Outpatient Process and Structural Measure Development and Maintenance Project at Mathematica Policy Research. With me today is

our project clinical lead and principal investigator, Dr. Tom Croghan.

Our project is tasked with developing and maintaining clinical quality measures supporting five of CMS's hospital quality reporting programs. As indicated in our project title, our focus is limited to development and maintenance of process and structural measures that define quality care for inpatient, outpatient, ambulatory surgical center, and cancer hospital patients. We also develop and maintain the electronic clinical quality measures, or eCQMs, for the hospital side of the meaningful use program.

The two measures we present and will be discussed today were developed by another CMS contractor, FMQAI. FMQAI's contract term ended prior to the initiation of this consensus development project. They had completed all measure development and testing and developed the initial drafts of the submission documents you reviewed.

We are here today representing CMS's interests in the development and, hopefully, endorsement of these measures. We want to acknowledge and thank FMQAI for their work and their generosity in helping us to understand this work and prepare for this presentation.

We do understand that the Committee may have questions about the measures that we are not able to answer as a result of our not being the original developer. We will note these questions, and we'll work with CMS and FMQAI to obtain answers for you to help inform your decision-making.

The measure under consideration now, timely evaluation of high-risk individuals in the emergency department, looks at the median time from ED arrival to qualified provider evaluation for patients triaged at the two highest severity levels on a five-level triage scale and addresses an important patient safety issue. This is an electronic measure data-sourced from the EHR.

Recent reports indicate that mean ED

wait times are rising. This is seen most critically for patients triaged at the two highest triage levels. Although waiting longer, patients triaged at the three lowest levels are, on average, obtaining care within the time frames recommended by the National Center for Health Statistics. Those at the highest two levels, though, are not.

In 2009, estimated mean wait times for those triaged at the highest or immediate level were 29 minutes, while recommended wait time is less than one minute. At the next highest level, emergent, the estimated mean wait times were 51 minutes while recommended wait times were between 1 and 14 minutes. Delay puts patients, especially those in most need of immediate attention and care, at risk.

FMQAI undertook extensive field testing at seven geographically and characteristically diverse hospitals. Reliability tests indicated ability to distinguish that the performance of at least one

hospital was statistically different from that of other hospitals. Construct validity tests identified cases where outcomes could have improved if care had not been delayed.

Feasibility testing demonstrated that all data elements were found to be available in the EHR systems and used by the hospitals, which included Epic, Cerner, and McKesson products. Criterion validity tests showed strong agreement between electronic and manual abstraction for two of the three data elements, ED arrival time and triage score, but less so for first provider contact time. These findings were reviewed by FMQAI's TEP and hospitals who acknowledged the challenge to accurately record this in the EHR but also noted an interest in increasing accuracy for their own quality improvement purposes.

This measure is not yet in a CMS program but has been reviewed and approved for continued development by the MAP for hospital inpatient quality reporting and the meaningful use programs. CMS has six other median time

measures in the hospital inpatient and hospital outpatient quality reporting programs. OP-20 is the closest similar measure, door-to-diagnostic evaluation by qualified medical personnel, but it reports median time to provider contact is chart abstracted and does not look at the severity level of the patient.

We thank you for your consideration and look forward to your review.

CO-CHAIR THRAEN: Kendall?

DR. WEBB: Okay. So I'm not going to repeat the introduction of the measure. I think it was adequately introduced. I'll go straight into evidence, if everybody is okay with that.

This is a process measure. There was a systematic review and QQC presented. I believe when I used the algorithm, actually, I came up with a high value for this.

To go over it just a little bit, being an ED doc and also an IT, this is probably perfect, actually, for me. Certainly, if you go over the data, you definitely want to triage your

highest-priority patients first. ESI is a triage system used in a lot of places but not everywhere, used mostly in larger places. And, actually, one of the things I'm going to want to talk about later is that it can actually be an onerous construct on some of the smaller EDs.

The ESI level-1 is basically the patient comes in dying or dead and you need to take care of them right away, and ESI 2 is they're not quite to that point but they'll get there if you don't take care of them right away, and that's why the 1-minute and 14-minute wait times are named here.

This does affect a large number of patients. It says here 130 million in 2010 was the number. And there are, you know, multiple studies to talk about, you know, getting to the patient faster is going to improve outcomes.

So this is in alignment with national priorities. This is electronically kept and can be gotten from there.

So just from an evidence perspective,

personally, when I looked through the algorithm, I thought the evidence was actually pretty high.

CO-CHAIR THRAEN: I have one clarification question. I thought I heard you say this was an eMeasure but in the documentation it says it is not. Could you clarify?

MS. CULLEN: It is an eMeasure --

CO-CHAIR THRAEN: It is an eMeasure.

MS. CULLEN: -- and electronic measure specifications were submitted.

CO-CHAIR THRAEN: Okay. So the documentation is a bit long. Jason, did you want to make any comments about, as an eMeasure, related to the evidence, is there anything?

MR. GOLDWATER: No, there's nothing on evidence. Their form was complete, and it's succinct.

CO-CHAIR THRAEN: Okay. Go ahead. Richard?

DR. BRILLI: I guess I just have one concern about the use of mean wait times for something like this. So mean wait times, even

median wait times but in particular mean wait times are highly subject to outliers. So you could have 99 patients who wait one minute and you have one patient who waits four and a half hours, and the meantime is totally skewed. So, to me, the measure ought to be percent of patients who achieve the goal. So if the goal is one minute or the goal is 14 minutes or whatever it ought to be, you want 99 percent or 90 percent of the patients or even 100 percent of the patients to achieve that goal.

I think mean wait times are very, it's a very bad measure to really figure out performance. It's what a lot of people do, but you really want the percent of patients who achieve the goal so that you take the outlier issue out of it in terms of time.

CO-CHAIR THRAEN: Any response?

MS. CULLEN: This measure is a median wait time, not a mean wait time.

DR. BRILLI: You have the same, you have the same issue, though. If you don't talk

about the number of patients who achieve the goal, you're still going to be affected by those, you know, the patients who have an excessive wait time.

DR. WEBB: So I have to tell you, as an ED doc, you're going to start out with this measure with a zero percent and you're not going to have any way to do any comparisons at a lot of hospitals because we're not meeting this measure right now. The vast majority of us are not meeting this measure right now.

DR. BRILLI: I get it. All I'm saying is that it's just going to be highly susceptible.

DR. WEBB: I understand. Yes, I understand.

CO-CHAIR SEPTIMUS: The percentage of patients who met that goal was about 25 percent, so it's not actually zero. And there is a correlation with the median wait time.

DR. ALEXANDER: For ED docs, 25 percent feels like zero.

DR. WEBB: And the number of hospitals

that were done where it looked at were seven hospitals, and none of them had a volume greater than 19,000.

DR. BRILLI: And just because it's not being done now or it's not -- it doesn't mean we should continue to perpetuate a measure that I think could be better if we used percentage to achieve as opposed to either median or mean wait times. This is an opportunity to change it.

CO-CHAIR THRAEN: So I think we need to go to Lisa and then Jason and then Josh.

MS. MCGIFFERT: Just a quick question on if it's an eMeasure, is the data collected automatically when somebody checks in and then when they get assessed, or how does that work?

MS. CULLEN: That depends upon the electronic record system. What was done in the testing was to compare what was recorded in the electronic record with someone going in and doing a manual abstraction to see if there were other keys that would identify that there might be potential other locations for that. And there

seemed to be, in some cases, good agreement, in some cases not good agreement.

CO-CHAIR THRAEN: Jason?

DR. ADELMAN: I recently had the opportunity to review a project that was proposing to have an improved system for classifying these different stages of emergency room triage, and the premise was that the current system was just often inaccurate with lots of variability. And so I'm not an expert in it. I'm just reflecting what I read.

But if that were true, then this measure, you know, will reflect the accuracy of triage, as opposed to the time, or it's capturing time of something that's very inaccurate. So I wonder if the developers and those who know more about this can comment on that.

CO-CHAIR THRAEN: Okay. That's a reliability question. Let's hold on that one. Is there any other questions related to the evidence? Go ahead, Josh and then Lillee.

DR. RISING: Hi. Josh Rising. I just

want to respond, you know, to the percent versus trying to use the mean versus median. I mean, I think the goal is to see all these patients as quickly as possible, not just necessarily, you know, within a 14-minute kind of window. So I do think kind of having the specific time, there is some value in that.

I guess one question that I did have, though, is that, you know, the data shows that, you know, there's a lot of large hospitals and presumably even more smaller hospitals that aren't using the ESI kind of scoring system at all. So I was curious how could the developers and others talk about, you know, the value of this measure if it's not even going to be able to be applied at a large number of facilities in the country?

DR. CROGHAN: The important concept with ESI is that it has five categories, and that's the valid reliable way and appropriate way of triage patients. The comparison is a three-level measure, and some of the ER docs in the

room may know this better. But those are not as reliable, and so there is room in this measure to use an alternative other than the ESI.

DR. RISING: Is the alternative system used at all the hospitals that aren't using the five-tier system?

DR. CROGHAN: I think most of the -- say again.

DR. RISING: So you said that, I said that it looks like there's a lot of hospitals that aren't using the five scoring system, and you said, well, an alternative is there's this other three-level system. But I wanted to know if the other hospitals then are using this alternative system.

DR. CROGHAN: I believe most of them are. I'm going to look at Dr. Pines here.

DR. PINES: Sure. So if you look across all hospitals, ESI is the most commonly used system. There are -- ESI is a resource-based system based on how many, how many resources a patient is going to need to be cared

for in the emergency department, and on the higher end it's about severity of illness. So ESI has shown to be more reliable than when you look at two nurses looking at the same patient comparing ESI versus some sort of time-based triage, you know, does this person need to be seen in 15 minutes, 30 minutes, ESI tends to be much better than that in terms of reliability. So it is the most reliable and most used measure but not used everywhere.

DR. CROGHAN: I think the correlation between the three- and the five-category systems, the five-category system is much more predictive of eventual hospitalization and mortality relative to the three-category system, as well. So it is, you know, implicitly, you're sort of driving people to use a five-categoric system.

DR. PINES: Yes, I recognize it sounds like it's the right system to you. It's just a challenge if there's a ton of facilities that aren't using it currently.

CO-CHAIR THRAEN: And I think that

speaks to usability. Again, we're at the evidence state. So if the cards that are up, do you have questions regarding the state of the evidence?

MS. ARDIZZONE: This is -- I just wanted to support the evidence. I'm saying that the evidence is extremely clear that this is beneficial, helpful. And, remember, they're just looking at ES-1s and 2s, so we're looking at the most critical patients, making sure that we're meeting the target for them.

CO-CHAIR THRAEN: Steve?

DR. LAWLESS: Yes, and the evidence, just a clarification of the evidence. In the justification of where the outcomes improve, the numbers that were used were cited, you know. Going from 77 minutes to 66 minutes were dramatic. The data that shows where we currently are with, you know, the 95th percentile is in the 12- to 14- to 16-minute range. So is the evidence matching to what you actually surveyed in the hospitals to what the literature says

where the improvement can be?

DR. CROGHAN: I'm not sure that, I mean, the seven hospitals were chosen conveniently because they had good data. And my guess is that it's not a representative sample and probably wouldn't reflect the national surveys done by CDC.

DR. LAWLESS: Right. So I'm asking -- I got that. I mean, they didn't have the 77 minutes waiting for the ES-1. But would you see, does the evidence support that, at the level where these hospitals are, at a lower level where these seven were randomly selected would have the same improvement and outcome?

DR. CROGHAN: I'm just going to make a point of order. Is that a validity question?

CO-CHAIR THRAEN: Steve, re-state it.

DR. LAWLESS: Is the evidence you're showing for justification that this is, does the evidence show that, at the level where you surveyed your hospitals, you would have the same outcome impact versus --

DR. CROGHAN: So if you're at 20 minutes or 14 minutes or whatever, you know, by implementing this, would you improve care?

DR. LAWLESS: Right.

CO-CHAIR THRAEN: Yes, that's a generalization question, so, yes, I think that falls into validity. Sorry. You'll have to answer it anyway.

DR. CROGHAN: I have to answer it.

CO-CHAIR THRAEN: Yes, but we'll give you a minute to think about it. Let's vote on the evidence.

MS. GELINAS: I just want quick yes/no. In the evidence, were there studies about the competency of the nurse being able to perform triage?

DR. CROGHAN: Yes.

MS. GELINAS: Because we found the same thing.

CO-CHAIR THRAEN: He answered yes.

MS. GELINAS: And it is a huge problem. We just standardized from a number of

ESI levels to a five-level system across multiple hospitals, multiple states. It's not easy. We have Cerner and Epic. It is not easy. Did the risk or adverse event landscape in your work in the past indicate through the evidence that, by applying the consistent levels and consistently measuring the adverse event and sentinel event, processes were improved?

DR. CROGHAN: We may have to get back to you on that one.

MS. GELINAS: Because I can help provide some of that, but I wanted to know if it's in the evidence base.

DR. CROGHAN: I'm not 100-percent sure I understand your question but --

CO-CHAIR THRAEN: Can we vote on the evidence?

MS. IBRAGIMOVA: Importance to measure and report, 1A, evidence structure, process, intermediate outcome. The votes are one high, only eligible if QQC submitted; two moderate; three low; four insufficient evidence.

CO-CHAIR THRAEN: Okay. We're missing a couple. We only have 19. We need one more. Victoria stepped out, okay. We're good.

MS. IBRAGIMOVA: So the results are 40 percent high, 55 percent moderate, 5 percent low, zero percent insufficient evidence.

CO-CHAIR THRAEN: Okay. Kendall, performance gap?

DR. WEBB: All right. So opportunity for improvement. I think we all see that there is opportunity for improvement here. As far as there were definitely, in these seven hospitals that they looked at there was definitely a wide variety of average minutes to get to the ES-1 and ES-2 levels. Again, I would note that the sample size was a convenient sample. And 19,000 visits per year is a small hospital for most. So whether it's generalizable to all hospitals nationwide is unknown, although I would actually think you would even, personally, I would think you would get a bigger gap if you looked at the large hospitals.

There was also some age and race disparity noted in some of the evidence. So as far as opportunity for improvement, I would say absolutely there's opportunity for improvement here.

CO-CHAIR THRAEN: Charlotte?

DR. ALEXANDER: I would like to make a plea to add language to your race disparity data.

CO-CHAIR THRAEN: Any other discussion or questions? Let's vote.

MS. IBRAGIMOVA: Importance to measure and report, 1B, performance gap, the votes are one high, two moderate, three low, four insufficient.

MS. THEBERGE: Ann, we need your vote.

MS. IBRAGIMOVA: The results are 64 percent high, 36 percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR THRAEN: Reliability?

DR. WEBB: Reliability -- all right, so for reliability, this is where we get into a

little bit of issue, or a lot of issue as far as I'm concerned.

It's noted by me and it was noted by the pre-committee members that there was poor agreement for the time to provider evaluated by the patient and what was documented in the chart. So - right, so what is that time point? And I think we're struggling with that nationwide actually.

Anecdotally, I can tell you that when they looked at their seven hospitals, they had one hospital that basically met these criteria and six that did not, and most likely that one that did had a practitioner out in the triage area would be the most likely cause for that, and the six that didn't did not have that practitioner out there.

That's not going to be a feasible model for a lot of hospitals for a lot of reasons to have a practitioner out in triage. But as far as meeting it, meeting a one-minute mark for ESI-1, otherwise it's going to be very difficult,

especially in the age of the electronic medical record.

So unless the doctor happens to be standing where the patient comes in the door, you're not going to meet that measure, or the PA or whatever, but you're not going to have a PA standing there to meet a patient who is coding.

Anecdotally, this has been a big issue in my hospital as well. We have instituted a thing called an MEI which is where we document what time we saw the patient, but it's really an attestation, right?

And a lot of -- I work in three hospitals, and all three hospitals have this where you go back and you say, I saw this patient at this time, right? It's an extra step you have to take. It's an extra page you have to fill out, but the medical record creates this artificial time line, right?

So if a patient comes in coding, I'm not going to click a button that says, okay, I'm seeing this patient now. I'm going to go see the

patient and then I'm going to go back and say, I saw this patient at this time.

And frankly, my MEIs, because of the way our process works and because I do go stand out in triage, I see the patient before the triage nurse gets done, and so my MEI actually occurs before arrival time.

So I see the patient before the patient even arrives in the ED, sometimes by up to 10 minutes. So I have a lot of concern about the data point, frankly --- anecdotally and as it's described in this documentation.

CO-CHAIR THRAEN: Questions?

DR. ADELMAN: I had asked about just the reliability of the triage system itself, not the time so much. And as I was waiting, I was doing a literature search and found that, you know, some articles that question it, because I read it in a proposal, but I don't have the references readily available.

But I thought maybe some of the experts would know because if the accuracy of the

measurement might not be as much about the time and the time intervals, but if too many, you know, threes are classified as fours and fours are classified as threes, that would undermine the reliability of the measure, so maybe you can respond to that.

CO-CHAIR THRAEN: But I think Laura's point that she made earlier, that it's really focused on the ones and the twos. Is that correct? Did I understand that correctly?

DR. ADELMAN: So then if twos are confused as threes, and threes are confused as twos.

CO-CHAIR THRAEN: Okay, Kendall?

DR. WEBB: And frequently twos change to threes or threes change to twos or ones --

CO-CHAIR THRAEN: Sure.

DR. WEBB: -- as part of the process.

CO-CHAIR THRAEN: Sure, any other questions before we vote? Let's vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties 2a

reliability, the votes are one high, two moderate, three low, four insufficient.

CO-CHAIR THRAEN: Does not pass.

MS. IBRAGIMOVA: So the results are zero percent high, 23 percent moderate, 59 percent low, 18 percent insufficient.

CO-CHAIR THRAEN: Is that it? Okay, sorry. So this does not pass the reliability test.

CO-CHAIR SEPTIMUS: Thank you very much. Thank you. The next measure is 0687, percent of residents. And I'm sorry we're going to go through the break. If there are those of you who need to take a break, just please do, but we're almost back on schedule here.

0687, percent of residents who are physically restrained, long stay, CMS. So we have the developers who I know are here. They were just dying to come back today. We have a new person though. So, do you want to -- or do you want to introduce yourself?

DR. SMITH: Good morning, I'm Laura

Smith from RTI. I am the nursing home lead on the CMS symptom management contract, and Nate Breg is with me today, and he will be doing the introduction for this measure.

MR. BREG: Thank you. The physical restraint of nursing facility residents is a safety concern discouraged by clinical experts and its prevention is a CMS priority.

For years, this MDS-based quality measure has reported to residents and families the rates of residents physically restrained, promoting patient safety.

The assessment items that determine this measure are valid and reliable, and the measure itself differentiates between facilities, and is stable and valid.

The measure shows low prevalence of restraint used, but it is important to maintain this measure to continue to discourage the practice, and to close racial and ethnic disparities.

The measure reports the percentage of

all long stay residents who are physically restrained daily during the seven days prior to the target MDS 3.0 assessment during their episode of nursing home care ending in the target quarter.

This measure addresses a CMS quality strategy priority. Use of physical restraints is associated with adverse physical and mental health outcomes.

The prevalence of physical restraint use is low and falling. The mean facility levels for this measure were 1.2 percent in quarter two 2014 and the median was zero. Two-thirds of facilities have perfect scores of zero.

The mean scores had decreased since quarter one 2011, which indicates continued quality improvement. Overall, studies show the measure and its items are reliable. Prior studies show that the restraints' items have good inter-rater reliability.

The quality measure itself is stable and has a noise-to-signal ratio of about 84

percent. Sixty-six percent of facilities have scores that differ from the mean, so high and low quality facilities can be distinguished using this measure.

There is good reason to believe this measure is valid. Less than 0.1 percent of all long stay episodes had missing data on restraint or related items. The quality measure shows no seasonality. This measure has strong face validity as it captures use of devices, clinical guidelines, and experts recognize as dangerous to nursing home residents.

This measure has been providing patients and their families with valuable information about the safety of nursing facilities through the CMS Nursing Home Compare website and the CMS five star rating system.

With quality measures on chemical restraints publicly reported by CMS, both physical and chemical restraints are discouraged. One-third of nursing facilities have some rate of daily restraint use, and there is evidence of

differences in restraint use across races and ethnicities of residents.

Furthermore, with the rates of falls among residents being publicly reported, it is important to report restraint use in order to discourage this as an anti-fall strategy. This quality measure remains important to keeping nursing facility residents safe.

CO-CHAIR SEPTIMUS: Okay, Kim, are you on the line? You're still on the line I hope, Kim.

DR. APPLEGATE: Yes, I am.

CO-CHAIR SEPTIMUS: Okay, Kim is the discussant and we'll go through the evidence, Kim.

DR. APPLEGATE: Okay, so this is 67 percent of residents who are physically restrained in long stay. Please note that the measure is complementary, so it's related but not competing with a couple of other measures, including an acute stay measure and also a related measure that captures a subset of

restraint measures, which I think is complementary but not competing.

So this measure reports the percentage of all long stay residents, defined as 101 day or longer stays, who were physically restrained daily during the seven days prior to the target of the Minimum Data Set 3.0 assessment.

So this is required data captured, and these are quarterly measures over a three-month period and at nursing facilities around the U.S.

I agree with the developers and thank them for the work they have done to look at trend data, and I think this is a process measure and so it has both the important positives but also some of the limitations of a process measure.

But I think that the evidence is strong in what they've shown of the value of continuing this measure. They've done an excellent summary of literature.

What I would ask is if the developers would consider a table of the many factors that they have summarized in the document of the

factors associated with increased and decreased restraint use rather than having summaries of the literature.

Because I found it fascinating, but it was very difficult to capture all of the factors, not consult but associations between increased and decreased restraint use that becomes somewhat circular in understanding whether they were truly associated or confounding.

You know, for example, some of the factors may be indirectly related to the increased or decreased restraint use and may be simply socioeconomic factors. So basically, I would summarize saying that this is a very, very important process measure to continue using it.

Should I stop there or continue?

CO-CHAIR SEPTIMUS: No, we're just going to talk about the evidence, Kim, thank you. Yanling?

DR. YU: Thank you, Yanling Yu. I just have a question to the developer. Have you compared the evidence of chemical restraint

versus physical restraint in any of those long term care facilities, because my understanding is the chemical restraint is pretty prevalent.

MR. BREG: The use of chemical restraints is more prevalent, but as part of the review of this measure, we didn't compare literature finding relationships between antipsychotics and health outcomes and comparing those relationships to those between physical restraints and health outcomes. We didn't compare the relative health outcomes of those two different types of restraints.

DR. YU: So do you have a plan in the future that would consider that, to monitor those types of restraints, because those can have a very serious side effect, you know, on the elderly?

DR. SMITH: So there is a paper that came out within the last year or two by Tamara Konetzka which looks at trends over time for physical restraints and over the same time period for chemical restraints, and you do see with the

advent of physical restraints being publicly reported that the rates of antipsychotics have been going up.

I can't remember, has there been a follow-up? So CMS is now publicly reporting two different chemical restraints, antipsychotic measures. I'm not sure that someone has done a look at kind of whether now we're seeing the trend lines do something different, but it is certainly something that we're aware of.

And when Tara McMullen was here yesterday from CMS, we have been talking a little bit about looking at some things like that and the interrelationships between those measures.

CO-CHAIR SEPTIMUS: Theresa?

MS. EDELSTEIN: Thanks, I just want to comment also on the same question. CMS has had a national initiative going on for the last couple of years to reduce the use of antipsychotic medications in nursing home residents, particularly those who have dementia, and the trends are going down.

The goal was achieved over the first few years to achieve an overall national 15 percent reduction in the use of antipsychotic medications in nursing home residents, and a new goal was set for an additional 25 percent reduction over, I think it's the next two or three years.

So there's good recognition that lowering physical restraint use had increased antipsychotic medication use and is now being addressed nationally, so there is progress being made.

CO-CHAIR SEPTIMUS: That's an excellent point. Kim, was that you on the phone? You were going to comment?

DR. APPLEGATE: Yes, I sort of raised my hand on the chat. Yes, I wanted to say that the developers were aware of that and they did talk about the CMS initiative and also the association with chemical restraints, and the increased use of the antipsychotic medication in the elderly associated with decreased use of

restraints, and the potential unintended consequences of, you know, the public reporting of the use of restraints.

So there is some data on that, and some of the references were included in this report. And I think that it's very important to also understand that there is a black box warning by the FDA for all antipsychotic medications because of the risk of arrhythmic deaths in the geriatric population, so there is an understanding that there is a downside to the chemical restraints.

CO-CHAIR SEPTIMUS: Pat?

DR. APPLEGATE: And I also -- if that's not something that the developers did discuss about the -- they have no detail about the serious adverse effects of the antipsychotic medications.

CO-CHAIR SEPTIMUS: Pat?

DR. QUIGLEY: Thank you, Pat Quigley. And I'd like to speak in support of keeping this a very clear indicator specific to physical

restraints. There is also an NQF measure for physical restraint reduction in acute care.

And the physical restraint issue is really a balancing measure in relationship to falls, because when we know there's been a long history or documented evidence that there is a negative correlation, when you saw the fall rates going down the restraint rates were going up, and it was physical restraints and restraint mobility.

The issue with chemical restraints is an issue of validity, being able to describe are we providing some of these medications to be able to indeed manage very difficult behavior versus limit mobility? And this is really focused on limiting mobility, this issue of restraints.

So I'd like to thank the presenter and the developer too for that excellent overview of the evidence surrounding this.

CO-CHAIR SEPTIMUS: Okay, I think we're ready to vote on the evidence.

DR. APPLEGATE: Ed, this is Kimberly.

I have one other point to make that the developers stated. The key point there really comes out with the evidence, and I'm sorry I didn't state this up front, is that restraints do not prevent major adverse events.

Although falls will increase when restraints and their abuse decreases, serious falls do not increase. So I think that's a really key point that is brought out with the report and the evidence.

DR. QUIGLEY: Excuse me, Dr. Septimus, if I may clarify, the issue with restraints and with falls is that if a patient falls with a restraint on, the severity of injury is greater, and that's evidence that's been published for years. So it's not the fall, it's the severity of injury. Thank you.

CO-CHAIR SEPTIMUS: Thanks for the clarification, okay.

DR. APPLEGATE: Could you ask the developers to clarify what their evidence is that they bring out because they say the opposite,

that serious falls do not increase with the decreased use of restraints?

CO-CHAIR SEPTIMUS: Pat, you'll have to put your mic on.

DR. QUIGLEY: Excuse me, I had mentioned with restraints. If a patient falls with a restraint on, for example from a bed, the severity of injury is greater.

DR. APPLEGATE: Okay, good. Okay, I was just making sure that we were agreeing. Thank you.

CO-CHAIR SEPTIMUS: Okay, now we can vote on the evidence.

MS. IBRAGIMOVA: Importance to measure and report 1a evidence structure of process intermediate outcomes. The votes are one high, only eligible if QQC submitted, two moderate, three low, four insufficient evidence.

And the results are 59 percent high, 41 percent moderate, zero percent low, zero percent insufficient evidence.

CO-CHAIR SEPTIMUS: Okay, now we go to

gap. Kimberly?

DR. APPLEGATE: I'm sorry about that. Let me just pull this up. They talk about the preponderance gaps, that there was a permanent and prevalence of difficult restraint use from the initial time, and they showed trends in the appendix.

The development listed only 66.9 percent utility of that perfect score, so I think there is appropriate evidence for potential improvement.

And then in terms of disparities, I think also they adequately discussed the gaps with -- basically the significant difference between black, Hispanic, and Medicaid patients in the appendix.

CO-CHAIR SEPTIMUS: Kimberly, are you finished, Kimberly? I just want to make sure I don't cut you off.

DR. APPLEGATE: Yes.

CO-CHAIR SEPTIMUS: Okay, Chris?

DR. COOK: Yes, I was looking in

further past the 66 --

CO-CHAIR SEPTIMUS: Chris, is your mic on?

DR. COOK: Yes, this is Chris Cook. Looking at the perfect scores of 66.9 percent then caused me to look further down into the, you know, the percentile ranks. And all the way through the 60th percentile with zero percent, 70th percentile is 0.9 percent, 80th percent 1.9 percent, 90th percentile 3.6, and that's all presented on page 36 and 37.

So the question to the developers is within that, can you statistically tell a difference between facilities with such a small difference? And I guess the question is, is the gap still there that allows for this measure to continue forward or has this been topped out?

MR. BREG: Sure, thank you for your question. The statistics that we present in the reliability section show that you are able to distinguish between facilities using this measure.

The signal-to-noise ratio is 0.84, which is acceptable for a facility level quality measure. And we also present stratified means that show that 66.4 percent of facilities had scores that were statistically significant from the mean at the 95 percent confidence interval.

Furthermore, a measure, as I understand the NQF guidance documentation, wouldn't be considered topped out if there is evidence of disparities. And as we show, there is evidence of racial and ethnic disparities.

And in the evidence documentation we submitted, there is one study that used logistic regression to show that there are different odds ratios for white and black residents for being restrained.

CO-CHAIR SEPTIMUS: Missy?

MS. DANFORTH: Yes, can you talk about why the mean reported on the Nursing Home Compare is so much higher than the mean in your measurement documents? You're showing a mean of 1.2 percent. They're showing a mean of something

like 20 percent, a national mean for the purposes of comparison.

So according to the data, the national mean is like much lower than what's being shown on Nursing Home Compare. Can you say why that is?

DR. SMITH: And what was the number?

MS. DANFORTH: Nursing Home Compare is showing a mean of -- hold on, I just had it in front of me. I'm sorry, give me one second.

DR. SMITH: I'm only asking just because that sounded very high.

MS. DANFORTH: Yes, so -- oh, I'm sorry. Okay, so if the mean is -- back to the question. They're showing a mean of 1.1, and then you're showing a mean of 1.2. So because the performance gap seems to be within the disparities piece, can you just sort of talk about how you're addressing that?

So if the performance gap seems to be with the disparities, which seems clear, but that seems to be washed away when you look at the

performance within the individual nursing homes. Can you just talk a little bit about how you're addressing that particular gap?

DR. SMITH: So I think one of the other issues to consider is what the rate should be for this measure. And since we're talking about daily restraints and -- one would argue that if you're having restraints every day, that means that's sort of the underlying cause of why the individual might be being restrained is not being addressed.

I think that the -- it seems like the ideal should actually be heading toward zero, and so I think we would still argue that the fact that we do have some facilities that have daily restraint use going on that there's still room for improvements.

You're correct that the way that the data is being reported, it's not obvious, but there is that racial disparity, but that there's still value in tracking this measure.

It's not only a publicly reported

measure but a measure that's used through the -- via the CASPER reporting system that the nursing homes use for their own internal purposes as well, and so it may be more of making the nursing homes aware of this being a particular issue than they may have a disparity, sort of a formal populations issue.

CO-CHAIR SEPTIMUS: Okay, let's vote on gap.

CO-CHAIR THRAEN: I just want to clarify. The only way that this goes into reserve status if you believe it's been topped out is if it does not pass the performance gap. Is that correct?

(No audible response)

CO-CHAIR THRAEN: All right.

MS. IBRAGIMOVA: Importance to measure and report 1b performance gap. The votes are one high, two moderate, three low, four insufficient. All right, so the results are 27 percent high, 50 percent moderate, 23 percent low, and zero percent insufficient.

CO-CHAIR SEPTIMUS: Reliability, Kim?

DR. APPLEGATE: Okay, reliability, the -- so just to briefly talk about how the residents are counted, they are defined as residents, as I said, who have a stay of 101 days or more. There's a separate measure for acute care.

Data and reliability testing was pulled from the Nursing Home Minimum Data Set 3.0, two serious reviews, the development and validation of MDS 3.0 and the RTI. The testing was done at the facility and agency level.

They ran reliability analyses to test reliability of the data element levels. The restraint items included in this measure have kappa statistics for gold standard nursing, sorry, nurse to facility nurse agreement ranging from 0.746 to 0.844, so it was very high. Limit restraint to in bed, and limit restraint in chair or out of bed both had perfect agreement.

The gold standard nurse to gold standard in nursing ratings have perfect

agreement for all items included in the measure except for chair prevents to rising data, which had a kappa of 0.887, also high.

And then I wanted to talk about what the numerator and denominator exclusions were so people understood that. The numerator is the number of long-stay residents with a selected targeted minimum data sets, and they have a number of different definitions of different restraints there in the MDS, and it has to be daily for seven days within that quarter.

And then the denominator exclusions were resident excluded if the denominator had missing data in any of the responses relevant to the question, so any of the different restraint questions, and if the facility sample had fewer than 30 residents in the facility, so if this was a small nursing home.

And then as the developer mentioned, they have a high --- signal-to-noise quality measure. So I think that from the reliability standpoint, it had a high rating.

CO-CHAIR SEPTIMUS: Okay, thank you. Let's vote on reliability.

MS. IBRAGIMOVA: Scientific acceptability of measure properties 2a reliability the votes are one high, two moderate, three low, and four insufficient. The results are 67 percent high, 33 percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR SEPTIMUS: Okay, are you ready for validity? Kimberly, any major points on validity?

DR. APPLEGATE: Yes, so the validity testing was both at the data element level and at the measure score level, and the RAND validation of the MDS 3.0 tested the criterion validity by comparing how different nurses assess the same residents.

Using the MDS 3.0, they compared gold standard research nurses to gold standard nurses, and they compared gold standard nurses to staff nurses trained by the gold standard nurses, and they used kappa statistics.

The restraint items included in this measure had kappa statistics for the gold standard nurse to facility nurse agreement ranging from 0.746 to 0.844, so high. The limb restraint in bed and limb restraint in chair or out of bed both had perfect agreement.

The other thing we were asked to discuss is the threats to validity, and I think they did two important analyses on the data which addressed this, and I don't see any threats to validity.

One was the correlation between this measure and the indirect measure 0674, and they looked at the percent of residents experiencing one or more falls with major injury in quarter three of 2013 and found that there was weak or not significant correlation. They had an R of -- 0.0145.

The other thing they did is they found a lack of evidence for a relationship between restraints and falls. It's possibly because the prevalence of physical restraint use and

incidence of falls with major injury are both very low, and they put in percent in parentheses, 1.2 percent and 1.6 percent respectively.

So they do raise that possibility that they might have a lack of evidence and raise that as a problem, a potential problem with validity which was talked about by the developers.

CO-CHAIR SEPTIMUS: Comments? Okay, let's vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties 2b validity, the votes are one high, two moderate, three low, four insufficient. And the results are 41 percent high, 55 percent moderate, five percent low, zero percent insufficient.

CO-CHAIR SEPTIMUS: Okay, feasibility?

DR. APPLEGATE: Oh, there was no concern with feasibility I don't think because even though it's not considered an eMeasure, all data elements are defined fields in electronic clinical data with the MDS. Any comments?

CO-CHAIR SEPTIMUS: No, let's vote.

DR. APPLEGATE: Okay.

MS. IBRAGIMOVA: So the feasibility the votes are one high, two moderate, three low, four insufficient. The results are 90 percent high, 10 percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR SEPTIMUS: Next is usability.

DR. APPLEGATE: Ed, can you remind me what --- if there are -- I mean, I did not see any particular concern about usability.

I just wanted to ask the developers a question about whether they had looked into one issue that I didn't see which was literature on the association between use of restraints and turnover rate of staff in terms of violent patients and violence against the staff.

There was no discussion of that, and I know that this has been an issue that, you know, comes up, you know, with why restraints are used and maybe under reporting of restraints, so I just wanted to hear a comment from the developers.

MR. BREG: We didn't come across any literature on violence against staff members. But I'd just like to repeat that this quality measure is -- reflects daily restraint use.

And as Dr. Smith pointed out earlier, it's the responsibility of the clinicians to identify what underlying factors lead to the behavior that leads to restraint use.

So if a facility is not -- is restraining a resident every day for seven days, there's a high probability that they're not performing the function of identifying what the underlying cause is.

And one clinical consultant to this project suggested that such a resident would probably be removed from the post-acute care facility if there was such a disturbing behavior. Does that answer your question?

DR. APPLEGATE: Sort of, but these are long-term stay patients.

MR. BREG: Yes.

DR. APPLEGATE: Well, I'm just asking

the question. Anyway, I think it's a highly useful measure, easily used and easily compared to others across facilities.

CO-CHAIR SEPTIMUS: And it's currently being reported, so --

DR. APPLEGATE: Right.

CO-CHAIR SEPTIMUS: Theresa?

MS. EDELSTEIN: Theresa Edelstein, just a quick comment to Kimberly's point. One of the things we see in nursing home populations, long-stay nursing home populations, is a growing number of folks who have had lifelong psychiatric diagnoses, and now they're living longer and making their way to the nursing home and have the physical needs as well as their psychiatric needs to be addressed.

So I just wonder whether we will continue to see improvement in this measure because of the changing nature of the population over time, and I think it's something that warrants some deeper exploration, whether that change is real, first of all, and what the impact

is on restraint use.

CO-CHAIR SEPTIMUS: Pat?

DR. QUIGLEY: Thank you, Pat Quigley. And I also wanted to respond to the question that Kim had presented, and thank you, Kim, for that. And I would suggest that in the patients you are describing that this would be the qualitative component of physical restraint use.

That patient or resident may need to have a restraint, but in policies and in long-term care, they would use the least restrictive restraint to still be able to reduce harm. Thank you.

CO-CHAIR SEPTIMUS: Okay, let's vote.

MS. IBRAGIMOVA: Usability and use the votes are one high, two moderate, three low, four insufficient. The results are 64 percent high, 36 percent moderate, zero percent low, zero percent insufficient information.

CO-CHAIR SEPTIMUS: Okay, and last one is it suitable for endorsement? So this is a yes or no.

MS. IBRAGIMOVA: Overall suitability for endorsement, does the measure meet and have criteria for endorsement? One yes, two no. The results are 100 percent yes, zero percent no.

DR. APPLEGATE: Excellent.

CO-CHAIR SEPTIMUS: Number five, fantastic. Okay, we'll tell you about some modifications in the schedule after the next one, but we are, as you can imagine, probably going to have to have a working lunch.

So the next one, we'll go in order, is 0689 Percent of Residents Who Lose Too Much Weight long-stay also from CMS. So, Laura, if you stay, do you have somebody else to work with you on this matter? Who is the discussant on this? Oh, it's Laura. There she is. Okay, Laura, remember efficiencies --- time, no pressure. Okay, developers? Introduce yourself.

DR. SMITH: Yes, so I'm joined by Dr. Qinghua Li also from RTI International. This is for the same contractor, CMS.

DR. LI: Okay, briefly, okay, this

measure requires the percentages of -- nursing from a resident to the target MDS assessment that indicates a weight loss of five percent among the last 30 days or 10 percent among the last six months, which is not a result of a physician prescribed weight loss regimen.

The data used for this measure is MDS 3, which is mandatory for all Medicare or Medicaid certified nursing homes. On this measure are strategies of serious strategy goal in alignment with one priority of the National Quality Strategy senior care.

It is important to maintain residents' nutritional status in nursing homes, and we have demonstrated that weight loss is the most objective and reproducible marker of nutritional status for nursing home residents.

Publicly reporting this measure will provide nursing homes incentive to monitor and maintain residents' weight and nutritional status. Since the last endorsement in 2011, two additional denominator exclusion criteria have

been applied.

First, prognosis of life expectancy of less than six months, second, receiving hospice care on the target assessment. Empirical evidence has shown that weight loss is a part of the trajectory for elderly with end stage diseases.

Weight maintenance or weight gain is not consistent with the goals of end of life care or patients' typical preferences at the end of life. To test whether these two additional exclusions are appropriate, RTI conducted analysis at the residents' facility and at measure levels.

The findings support excluding residents who are receiving hospice care or are having a less than six-month life expectancy. We also received public comments and a subject matter expert's input supporting these addition exclusions.

To reevaluate this measure, we used data on all eligible long-stay residents from our

Medicare and Medicaid certified nursing homes from 2011 to 2014.

Nationally, this measure is pretty stable with a small variation from 5.7 percent in quarter three 2014 to 6.8 percent in quarter one 2013. The MDS 3 data elements used to calculate the weight loss measure were demonstrated to be reliable.

The analysis on quarter measure reliability showed that while the facility scores on this measure were stable between two reporting periods, the facility ranks on this measure changed more frequently. Both the weight loss annually on the MDS 3.0 and the weight loss quarterly measure were demonstrated to have moderate to high validity.

We also examined the proportions of facility scores for this measure that are significantly different from the national facility level mean.

The findings indicated that this measure can successfully distinguish facilities

in which there are quality concerns related to weight loss from high quality nursing homes where residents' nutritional status is managed very well.

One possible unintended consequence is that the increase in use of a feeding tube or other aggressive feeding programs amongst some residents. However, there is no evidence indicating such increase in the feeding tube use or other aggressive feeding programs.

A recent review of quarterly data from quarter two 2012 to quarter four 2014 showed a slow, but a very steady decrease in the feeding tube use in nursing homes.

So in conclusion, this measure is very important, valid, and a reliable quantum measure for nursing homes. Thank you.

CO-CHAIR SEPTIMUS: So, Laura, this is an outcome measure.

MS. ARDIZZONE: Yes.

CO-CHAIR SEPTIMUS: And so anyway, go, Laura.

MS. ARDIZZONE: So this is re-endorsement 689, percentage of long-term stay nursing home residents with MDS assessment that indicates weight loss of five percent or more of the baseline weight in the last 30 days, or 10 percent or more of the baseline weight in the last six months, and this is not a result of a physician prescribed weight loss regimen.

I think the evidence is very strong. I think this is an outcome measure, something we keep asking for in this committee. So on evidence, I think the evidence is supportive of the measure focus, so I think the evidence is strong.

CO-CHAIR SEPTIMUS: Comments? Yes, Steve?

DR. LAWLESS: I know it's a technicality. It shouldn't be increased mortality. It should be decreased survival because everybody dies.

DR. SMITH: I'm sorry, are you talking --

DR. LAWLESS: Yes, just in terms of the developer. It really is not an increased mortality factor, it's a decreased survival.

And so from a technical standpoint, it's actually -- it's -- it shows a little bit -- it's a little bit of a different slant to it, but it also ---- when you publicly -- when you report it, it's not likely to die. It's how long less of a survival do you have by that, so for the developers maybe just to consider that.

CO-CHAIR SEPTIMUS: Okay, let's vote.

MS. IBRAGIMOVA: For importance to measure and report 1a, evidence health outcome or PRO, the votes are one yes, two no.

MR. ANDERSON: Kimberly, we don't have your vote.

MS. IBRAGIMOVA: The results are 90 percent yes, 10 percent no.

CO-CHAIR SEPTIMUS: Okay, next would be gap.

MS. ARDIZZONE: My only comment on gap, and I struggled with this a little bit, is

there's nothing to indicate any disparity data, either race or socioeconomic status differences, and there really has been no observed improvements since the original measure in 2011.

So while I think it's important, I think it's a marker of something, I think it's an outcome, there just hasn't been showing any ---- that there's an improvement, or ---- no change, so that would be my only comment.

CO-CHAIR SEPTIMUS: Developers want to comment on that? I think there is something about disparities in here, unless I misread it, but what about the other comment about improvement?

DR. LI: Thank you for your comments. We recognize this kind of no improvement indication in this measure, and we didn't find the racial or socioeconomic disparity in this measure either.

As we have mentioned, while maintaining residents' nutritional status is very important in nursing homes, weight loss has been

found to be the most objective and reproducible marker of nursing home residents' nutritional status.

And also, actually the lack of a change in this quality measure may reflect that nursing homes are not improving quality, meaning taking care of patients' nutritional status. We think this might further kind of highlight the importance of keeping published -- publicly reporting this measure.

So with this measure being kind of retired, if we stop publicly reporting this measure, nursing homes may lose this incentive to keep monitoring and maintaining residents' nutritional status.

CO-CHAIR SEPTIMUS: Theresa?

MS. EDELSTEIN: Theresa Edelstein. Respectfully, I would like to ask you to not jump to conclusions or make assumptions about whether or not nursing homes are focused enough on improving nutritional status.

It's very well documented that the age

and acuity of nursing home residents is increasing over time, especially as we work very hard to keep people at home and in their communities receiving long-term care for much longer periods of time.

So the frailty and acuity of the nursing home population is increasing significantly, and with that comes difficulty in maintaining nutritional status.

CO-CHAIR SEPTIMUS: Okay, let's vote on gap.

MS. IBRAGIMOVA: Importance to measure and report 1b, performance gap. The votes are one high, two moderate, three low, four insufficient.

The results are 29 percent high, 57 percent moderate, 10 percent low, five percent insufficient.

CO-CHAIR SEPTIMUS: Reliability, Laura?

MS. ARDIZZONE: Okay, so they presented two separate tests for reliability. So

they ran -- did some testing on the MDS and they found really high kappa scores on data element.

There was a lot of discussion on the RTI analysis on performance that the stability analysis, the signal-to-noise was low, that maybe this measure isn't particularly reliable in separating facility characteristics from the noise of the population.

But I think you did a pretty good assessment at saying that the analysis on quality measures was high, but the facility scores were stable between two consecutive reporting periods, but the facility ranks may change frequently, and that might be associated with a relatively narrow range of the measure.

So I think you explained it. I don't think -- I think it's high on the data elements. I think at the facility level it may not be as high, if you wanted to comment on that.

CO-CHAIR SEPTIMUS: Jason?

DR. ADELMAN: I struggled with the reliability on this one because inherent to me

that both the numerator and the denominator has things that are -- would be very hard to measure reliably.

Like in the numerator, a weight change of five percent, which if you weigh 150 pounds, would be seven, seven-and-a-half pounds. Like in the hospital we see weights, you know, go up and down all the time like that.

And in the denominator, it has, I guess, a prediction of the life expectancy at six months. That seems very hard to reliably predict.

And so with both of these like -- so I understand that the kappa score was such, although the methods by which the reliability testing was done is not clear to me.

And inherently since it intuitively doesn't make sense that it would be super reliable, I was wondering if the developers could explain more about really how the reliability testing was done and how they addressed these variables.

DR. SMITH: So with regard to the item level reliability, that work was done using pairs of raters that have both been -- received training on how to complete the MDS data set. There was an agreed upon protocol for how that evaluation would be done for whatever set of items that there were.

I think they evaluate quite a few for that MDS data set, and basically were evaluating the person at the same time. So that rating is more about sort of the repeatability between the raters. You are referring more to this issue of the changeability of weight loss over time.

I guess I would argue that while we might not be capturing all of the instances where weight loss would have occurred of that degree, that we are sort of cross-sectionally going to be capturing weight loss and that there's not going to be one -- shouldn't be one particular bias being introduced into the measure because of that.

And certainly someone who might be

having that much of a change from moment to moment, this is a -- the assessments are done quarterly, but there should be, as a part of care, some monitoring -- a part of regular care, monitoring for weight loss.

And so the item is really asking about have they had this happen? So if they are having that much of a weight loss over this period of time, it should be reporting because that's an indicator of there being an issue. Okay, so that was the item level reliability.

You had asked about measure level reliability ---- oh, prognosis, sorry. And then the prognosis -- so that is -- again, we'll acknowledge that being a potential concern.

I guess, one of the things that we've talked about is sort of the importance of, regardless if it was not -- the reliability is not as high, it's still a very important group of individuals to be identifying in order to make sure that we're not putting them at risk for interventions that would be potentially -- really

against other objectives in terms of meeting preferences at the end of life.

Qinghua, do you want to add to that?

DR. LI: Actually for the prognosis item, which this is based on physicians' prognosis. This is from residents' medical record, which are kept by their physicians.

Also, the RAND report found that the kappa statistics for this specific item between the gold standard nurses was 0.87, and between the gold standard nurse and the facility nurse was 0.96, which means the reliability is very high.

CO-CHAIR SEPTIMUS: Pat?

DR. QUIGLEY: Thank you, Pat Quigley. My question in relationship to reliability, very similar to Dr. Adelman's, in that the numerator is those who have weight loss over a 30-day period of time versus a six-month period of time, and the RAI is completed quarterly.

So would it not be better to just go ahead and have the period of time of any weight

loss -- and I was interested in weight gain, because weight gain in older people in wheelchairs is very difficult and impairs mobility -- but is to have the measure just be related to the 90-day period in which the RAI is done, rather than the 30 days and then the six-month period which actually covers two RAI periods?

The RAI, for everyone, is a resident assessment interview that's done. It's the quarterly completion of the team review of the patient, because you have an "or" component. It's 30 days or six months.

DR. SMITH: Right, and I think the idea in having that potential shorter time period is to try to capture a precipitous loss or that variability that we were talking about a moment ago.

That perhaps if you compare a person's weight now to three months ago, there might not be that much change, but if you look at that 30-day period back, they may have had a weight

increase and then a weight loss in that time period, and what you're capturing with looking at that 30-day window is this variability in the weight that might be a concern.

DR. QUIGLEY: Yes, thank you, but the question is really that six-month period that's added to it because that covers two quarters and two RAIs.

DR. SMITH: So you're concerned about having the longer time period?

DR. QUIGLEY: Yes, yes.

DR. SMITH: Well, again, I think the idea is sort of you're getting a slightly different trajectory there that you may not have a major weight loss within a 30-day period.

But looking over the -- sorry, the larger ---- two time periods -- excuse me, for the weight loss may be revealing some sort of slower process of decline related to weight loss that you might not see on that shorter time period.

CO-CHAIR SEPTIMUS: Okay, let's vote

on reliability.

MS. IBRAGIMOVA: Scientific acceptability of measure properties 2a, reliability. The votes are one high, two moderate, three low, four insufficient.

And the results are four percent high, 70 percent moderate, 22 percent low, four percent insufficient.

CO-CHAIR SEPTIMUS: Validity, Laura?

MS. ARDIZZONE: For validity they did data element and performance score level testing. Both indicate acceptable validity testing.

They also did exclusion testing. There is no risk adjustment and the exclusions seem appropriate. So they exclude data if they don't have the right assessment type, if they have a six-month prognosis, or they're hospice patients, or if they are residents less than 30 days. CO-CHAIR SEPTIMUS: I have a question not related to the measure for my own education. Okay?

Why is weight loss most common in the

first quarter of the year?

DR. LI: This is an observation?

CO-CHAIR SEPTIMUS: Weight loss is most common in the first quarter of the year. I'm just curious, Theresa. I mean, I don't want to put you on the -- I'm asking for -- I actually read this stuff.

MS. EDELSTEIN: Well, it is right after the holiday season. It's also a time period when influenza is significant -- can be significant and other -- and respiratory illnesses and people don't eat when they don't feel well.

CO-CHAIR SEPTIMUS: I'm just curious.

Okay. Ready to vote on validity. Let's vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2(b), validity. The votes are one, high; two, moderate; three, low; four insufficient.

(Voting.)

MS. IBRAGIMOVA: This might take a

second.

(Comments off record.)

CO-CHAIR SEPTIMUS: All right. Do you want to vote by hand for this one while she -- why don't we do that.

Okay. So, all those who say high, raise your hand. And vote by via phone for -- high. All right. So, zero for high.

Okay. Moderate. Who wants to count?

(Pause.)

CO-CHAIR SEPTIMUS: And on the phone? Okay. So, one high and the rest are moderate. No one for low or insufficient.

Okay. So, now we go to feasibility.

MS. ARDIZZONE: I have no concerns about feasibility. I think it's important to continue using this measure.

CO-CHAIR SEPTIMUS: Okay. We'll vote by hand again. Is that okay with everybody?

Okay. So, feasibility. High.

MS. SPEAKER: Somebody needs to count.

CO-CHAIR SEPTIMUS: 14. Moderate.

(Pause.)

CO-CHAIR SEPTIMUS: Low. One low. Did the people on the phone vote? So, what did they vote? So, we have a medium and a high. Okay.

So, what was the final vote then?

MS. THEBERGE: What was the number for moderate?

MR. SPEAKER: Six.

MS. THEBERGE: Six, plus two on the phone?

CO-CHAIR SEPTIMUS: Six plus one is seven, and one additional one for high. Okay.

Usability. Laura.

MS. ARDIZZONE: This is already publicly reported. I think continued use of this measure can encourage adoption of improved processes, improved quality of care.

As we have more and more of an aging population, I think it's really important that we find quantitative markers of quality care for elders.

CO-CHAIR SEPTIMUS: Okay. I see no comments. We can once again vote by hand. High.

(Voting.)

CO-CHAIR SEPTIMUS: Drew, what are you doing?

(Comments off record.)

MS. IBRAGIMOVA: 20.

CO-CHAIR SEPTIMUS: All right. Moderate. And that's to 22. So, no low and no -- okay. Great.

Okay. Now, the measure is acceptable for public reporting. It's a yes or no. I'm getting to learn this thing. For endorsement. So, all those in favor of endorsement.

(Voting.)

CO-CHAIR SEPTIMUS: How about the two on the line?

MR. ANDERSON: Ann, could you -- oh, yes. Okay.

CO-CHAIR SEPTIMUS: You could put your hands down. We've got 22. So, another unanimous.

Okay. Fantastic. Thank you so much. So, this is what we're going to -- I'm going to turn this over to Iona who's going to take us hopefully most of the rest of the way.

And the next one we're going to go to falls, 0101. Thank you.

CO-CHAIR THRAEN: So, 0101, it's from yesterday's list. So, it's at the beginning of your agenda. And it's the Falls: Screening, Risk-Assessment and Plan of Care to Prevent Future Falls developed by the National Committee for Quality Assurance.

Could you please join us at the table?

(Pause.)

CO-CHAIR SEPTIMUS: And just in case you're wondering, lunch will be at 12:30. And we may have to make it a working lunch.

CO-CHAIR THRAEN: All right. Who's the speaker on this one? Pat.

MS. GIOVANNETTI: Yes.

CO-CHAIR THRAEN: Better be. So, developers, would you like to open your

statement?

MS. GIOVANNETTI: Yes.

CO-CHAIR THRAEN: Introduce yourselves as well.

DR. GIOVANNETTI: My name is Erin Giovannetti. I'm a research scientist with the National Committee for Quality Assurance. And I'm joined by Dr. Mary Barton who is our vice president for performance measurement.

So, thank you for allowing us to come today to present this measure, which I know is a little bit out of sequence.

This measure is a maintenance endorsement of a measure that previously came before a similar panel in very late 2012. It looks at falls risk prevention in the older adult population.

It is three indicators that are really all meant to be reported together as a group, because they look at a continuum of care for falls.

It starts by a measure that looks

first at whether or not you screened for whether or not you have a population at risk of falls. And this is really just asking -- assessing whether or not older adults have had a fall in the past year, two or more falls or a fall with injury, which then defines the denominator for the next two measures, which get into how do you prevent future falls in this at-risk population.

So, looking at that population, the people who have two or more falls, or a fall with injury, did you do a multifactorial risk assessment?

And we define what -- there are many different factors you could look at. We define a minimum set.

And then, did you do a plan of care including consideration of Vitamin D therapy and exercise or physical therapy.

This is based off of evidence from the US Preventive Services Task Force and the American Geriatric Society.

The USPSTF has guidelines that put in

place that all older adults who have a history of falls should have consideration of Vitamin D therapy and exercise or physical therapy.

It's also based off of recommendations from both the AGS and USPSTF that all older adults should be asked about whether or not they have fallen.

And then the one area where the evidence gets a little bit more difficult to interpret is around risk assessment for falls.

This is an area where the USPSTF has given a C recommendation that they basically said that the evidence -- it's really difficult to identify what's the population where a multifactorial risk assessment has a significant benefit. They note that it has a small net benefit for -- across all the studies.

The American geriatric society interpreted the evidence slightly differently and they found that there was a significant net benefit for adults who have a history of falls. So, those adults which they define as two or more

falls, or a fall with injury.

So, that's the definition that we are using as the at-risk population who would benefit from a falls risk assessment.

This measure is specified for medical record review collection. So, unfortunately these are not things that are normally found in billing codes. So, this is required to look at medical records to determine whether or not people have been asked about falls, whether or not those people who are at risk of future falls have had a risk assessment and a plan of care has been put in place.

It is operationalized in the PQRS program in two different ways. One is in their claims and registry reporting option, which is an optional reporting program for providers. And then also in their GPRO program, which is the reporting for Medicare shared savings plans and group practices.

The testing that we did on this was done back in 2009 looking at medical record

reviews. So, that's the way that this measure is specified and being presented for endorsement.

CO-CHAIR THRAEN: Okay. Pat, do you want to start with the evidence?

DR. QUIGLEY: Pat Quigley, and I'd like to thank Erin for that excellent overview. And to also say that she did an excellent overview in terms of the evidence.

The evidence in supporting this from the American Geriatric Society is also endorsed by the British Geriatric Society and the American Organization of Orthopedic Surgeons.

So, the evidence that is presented to support each aspect of the three elements of this measure have been graded by the United States Preventive Services Task Force.

And you have that discussion on Page 10 of this report, but I would like to say that there's been extensive effort in trying to get this measure, the AGS guidelines into primary care.

And that actually started in 2010 --

excuse me -- 2001. Reemphasized in 2010 with an update of the evidence.

And I'd also like to add to your presentation of this measure in that this measure was originally presented and endorsed in 2007. So, I think you did a great job in terms of the evidence and I think we can go forward, but I did want to make sure that everyone does appreciate that this is specific to those people who have fallen more than once or had an injurious fall.

And that's the more vulnerable side of the algorithm. And that's why the United States Preventive Services Task Force decided not to just go with anyone who had a fall in the last year, but the people who are more vulnerable in the algorithm who have had more than one fall or an injurious fall that they need to be worked up.

CO-CHAIR THRAEN: All right. Lisa, then Ed.

MS. MCGIFFERT: Two questions. Who is the population of patients? What is the population of patients covered by this?

Inpatient? Outpatient? Nursing Home? Assisted living?

DR. QUIGLEY: It's ambulatory care. Ambulatory care.

MS. MCGIFFERT: So, somebody put in a comment to recommend removing nursing home and assisted living patients. Is that valid? Is that a valid comment?

DR. GIOVANNETTI: Sorry. So, this measure actually -- it is a measure of providers. And so, included in this are eligible providers who may make visits and provide care in nursing homes.

So, some of the codes that are used to determine the denominator, which is a visit with an eligible provider, include claims codes for visits to rest homes, nursing homes, non-acute inpatient settings.

And so, it is possible that, you know, providers choosing to report on this measure through the PQRS program that have those types of visits would be included.

It is not a measure that is used to evaluate an institution. So, it's not used to evaluate a nursing home. It's used for providers.

DR. QUIGLEY: Provider.

MS. MCGIFFERT: Okay. Thank you.

And this seems really similar to the one we heard yesterday, but -- they're very close, but this is assessing whether the provider does it rather than whether it happens in the facility.

DR. QUIGLEY: Yes, and this is structure or process. So, that's why we -- there is a component to ask about fall risk, and then complete a fall risk and get a plan of care.

So, it's not linked to an outcome and reduce in fall, the reduction in terms of reducing risk, it's the structure and the process.

MS. MCGIFFERT: That was the one -- what was it? Just trying to remember. It was the one I did. Can't remember the number.

Well, anyway, they had three parts. Same thing. Assessment, plan, all of that.

DR. QUIGLEY: Pressure ulcers. Wasn't it pressure ulcers?

MS. MCGIFFERT: It was pressure ulcers. Thank you.

DR. QUIGLEY: Oh, you're welcome.

CO-CHAIR SEPTIMUS: All right. Just a quick question. It says here on Page 1 of our worksheet that identify patients at risk and that family physicians have a pivotal role in screening older patients.

Is there a particular reason you mentioned family physicians? Shouldn't that be any? It could be internal medicine. It could be any primary care physician caring for these patients. So, why do you mention family physicians only here?

DR. GIOVANNETTI: It could -- you are absolutely correct. It could be any type of physician. And, in fact, the eligible providers that can report on this include the whole range

of providers.

I think the family physician specifically references -- the reference that that came from was specific to family providers, but you are correct.

CO-CHAIR THRAEN: Any other -- oh, Lisa, are you done? Any other questions about evidence? Let's vote.

MS. IBRAGIMOVA: It should be working. Importance to measure and report, 1(a), evidence structure process intermediate outcomes. Votes are one, high only eligible if QQC submitted; two, moderate; three, low; four, insufficient evidence.

(Voting.)

(Discussion off the record.)

MS. IBRAGIMOVA: The results are 59 percent high. 36 percent moderate. Five percent low. Zero percent insufficient information.

CO-CHAIR THRAEN: Performance gap, Pat.

DR. QUIGLEY: Thank you.

In relationship to the performance gap, there's a clear opportunity for improvement and to get this into the hands of -- the practice of providers.

This measure is reported on PQRS, the Physician Quality Reporting System. For the first measure which is Rate A, screening for future risk, the performance measure was 41.5 percent. So, there's more opportunity, as you can tell.

The rates range from 9.8, the tenth percentile, to 79.8 percent, the 90th percentile, of those that were eligible.

For the second component of this for risk assessment of falls for fall rate, this was 5.2 percent that chose to actually report this.

So, I think that this was an issue in relationship to choosing to report this in the PQRS.

For Rate 5, the plan of care is also used in PQRS which was 78.9 percent, which was of those who chose to report it. So, this was much

higher and actually having a plan of care in place.

And as was mentioned by the developer, this is to address balance, as well as Vitamin D and C.

So, with that there is opportunity. There's also discussion in our report on disparities. And the disparities in relationship to this are trying to look at those that are associated with repeat falls that are falling higher that have -- more than twice and what we need to do in terms of their care planning.

And also people over the age of 75 who are four to five times more likely to be falling even greater and having a serious injury.

There was disparities in older whites that were 2.7 times more likely to die as a result of a fall than African Americans. So, there is a race disparity here. So, there is great opportunity for improvement.

CO-CHAIR THRAEN: Victoria.

DR. RICH: (Speaking off mic)

particularly when it deals with disparities. When you say that there's limited stratified, is that something in the future to be looking at?

I think with the diversity of our populations it seems to be such a keen, important measure to consider.

DR. GIOVANNETTI: I wholeheartedly agree. And I think that CMS is definitely moving in the direction of trying to collect more information stratified by race and ethnicity, as well as things like language barriers, which I think would be very important.

DR. RICH: Yes. Thank you.

CO-CHAIR THRAEN: Well, you cut Charlotte off.

(Comments off record.)

CO-CHAIR THRAEN: All right. Any other questions for performance gap?

(No questions.)

CO-CHAIR THRAEN: We'll vote.

MS. IBRAGIMOVA: Importance to measure and report, 1(b), performance. The votes are

one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: The results are 65 percent high, 26 percent moderate, nine percent low, zero percent insufficient.

CO-CHAIR THRAEN: Okay. Reliability.

DR. QUIGLEY: Thank you. There's no issues with reliability. They're very clear that -- but I did want to mention that there is an exclusion in terms of the denominator; those who are non-ambulatory, those who are in wheelchairs.

CO-CHAIR THRAEN: Any questions?

(No questions.)

CO-CHAIR THRAEN: Vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2(a), reliability. Votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

CO-CHAIR THRAEN: Could you revote again?

(Revoting.)

MS. IBRAGIMOVA: The results are 48 percent high. 48 percent moderate. Four percent low. Zero percent insufficient.

CO-CHAIR THRAEN: All right. Validity.

DR. QUIGLEY: For validity, there was face validity that was utilized for this measure, this component of the measure.

The American Medical Association convened a physician consortium of physicians from multiple specialty areas.

They had a 23-member physician consortium and they rated the validity of each of the components of this measure on a five-point rating scale where one was strongly disagree, to five, strongly agreed. And all three of the components rated over 4.3. So, there was a high level of face validity.

CO-CHAIR THRAEN: Any questions?

(No questions.)

CO-CHAIR THRAEN: All right. Let's

vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2(b), validity. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: The results are 48 percent high. 52 percent moderate. Zero percent low. Zero percent insufficient.

CO-CHAIR THRAEN: All right. Feasibility is next, I think.

DR. QUIGLEY: Yes. With feasibility, this data is collected through the administrative claims data, electronic claims data and then also paper medical record.

So, that's an opportunity to deal with the issues surrounding feasibility and then also whether or not the providers are actually reporting it.

It's not an eMeasure, but the data collection provided by the developer indicates few concerns related to feasibility.

CO-CHAIR THRAEN: Questions.

(No questions.)

CO-CHAIR THRAEN: Vote.

MS. IBRAGIMOVA: Feasibility. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: The results are 35 percent high. 61 percent moderate. Four percent low. Zero percent insufficient.

CO-CHAIR THRAEN: Usability. Pat.

DR. QUIGLEY: Thank you.

In terms of usability this clearly identifies the opportunity for continuing to address the fall risk needs --- fall reduction needs in the elder population.

So, recognizing the opportunity in relationship to the gap, there's a lot of opportunity to be able to improve practice. And in addition to usability, it helps us to be able to appreciate that.

The reduction of fall risk really

requires more than balancing, gait, mobility and calcium, Vitamin D.

The AGS guidelines, BGS guidelines, American Orthopedic Surgery, there is a multifactorial approach and there's lots of opportunity to be able to reduce risk.

And this is to make -- help elder patients be healthier. It's not just about whether or not they fell. It's being able to reduce their risk. And that's the focus of these guidelines.

So, there's a lot of opportunity for usability of this if we could get this into the hands of providers.

CO-CHAIR THRAEN: Questions.

Yes. Go ahead.

DR. YU: I have two questions. One is about the comments about there have been little variation in performance across providers which was to report.

This tied to the previous statement that payment program, and said that the measure

is the currently used payment program.

So, I would just -- can you describe or explain it why -- how you tie to payment program when you have so little variation among the performance.

DR. GIOVANNETTI: So, the payment program that this is tied to, there's two different programs.

One is the PQRS voluntary reporting program in which it is a pay for reporting. So, you are not paid based off of your performance. You are paid because you've reported the measure and providers choose which measures they want to report on.

And so, what you end up with is providers that are doing well in a particular area, choose to report on that measure.

And so, in the two measures that look at risk assessment and plan of care we see high rates with not a huge amount of variation primarily because the providers choosing to report on them are doing those things and

documenting them.

The other one, the screening, is actually being used in the GPRO program, which is not voluntary. That one is for the Medicare shared savings programs and group practices are required to report on that.

And that is actually something where we're starting to see more of the pay for performance. And so, we see more variation in that measure.

There is movement of the PQRS program to move from a voluntary reporting program to become more of a mandatory reporting program with payment for performance as opposed to payment for reporting, but we're not there yet.

DR. YU: Yeah. Okay. Thanks for clarifying.

My other question is definitely you want to see more people or physicians, you know, to encourage them to report.

And there are comments that PQRS involved in the reporting, you know, increase is

expected, I just wondering what is the plan or do you have any, you know, explanation what is the improvement down the road that you would encourage more reporting on this type of thing

DR. GIOVANNETTI: So, NCQA does not run the PQRS program. That's run by the Centers for Medicare and Medicaid.

They have put forth several proposals for improving that program that actually many of which I believe were laid out in the -- some of them in the MACRA legislation. The SGR.

So, unfortunately, that's not a program that's within our control, but CMS is working towards more of a value-based purchasing model in which the measures that are in the PQRS program will be used in that way.

DR. YU: Okay. Thank you.

CO-CHAIR THRAEN: Laura.

MS. ARDIZZONE: Yes. Would you be able to comment on some of the -- there seems to be a lot of competing measures and there's actually one that really sounds similar, NQF

0035, Fall Risk Management for All Older Adults Across All Settings.

I'm concerned about the burden of reporting on, you know.

DR. GIOVANNETTI: I'd be happy to.

So, 0035 is a measure for health plans as opposed to providers. It does look at the similar concepts and we've actually -- we also stored that measure in CQA. And so, we harmonize it in terms of concepts, but that measure is actually collected through a survey, not through physician reporting.

So, if a physician is doing the things to meet this measure and the patient is understanding what's going on, ideally the patient will report on the survey that those things happened.

So, it's not adding to physician burden in terms of having to report on two different measures.

The reason that we collected for a survey for the health plan has to do with some of

these issues around the burden of data collection.

So, as we noted before, a lot of this information has yet to go to the medical record to find it.

And so, physician that may do that in the PQRS program we felt that for the health plan reporting rather than health plans having to go back out and do this again, this could be better captured through a survey.

The survey also gets some other elements, which is how much is the patient actually understanding that this has happened for them.

So, if they don't know that anyone has actually advised them on how to prevent falls, that's a kind of other end of the spectrum.

So, they tell you slightly different things about the same underlying measure concepts that are aligned.

CO-CHAIR THRAEN: Kimberly on the line, you had a comment?

DR. APPLEGATE: Oh, it was actually a comment about the PQRS program. My understanding is that it's actually a penalty program now in 2015 and it will pay out -- I mean, it will assign penalties in 2017 for the 2015 program reporting of those PQRS metrics. It's just a comment.

CO-CHAIR THRAEN: All right. Any other -- Laura, did you have a -- any other conversation, questions?

(No comments.)

CO-CHAIT THRAEN: Shall we vote?

MS. IBRAGIMOVA: Usability and use. The votes are one, high; two, moderate; three, low; four, insufficient information.

(Voting.)

MS. IBRAGIMOVA: The results are 17 percent high. 74 percent moderate. Nine percent low. Zero percent insufficient information.

CO-CHAIR THRAEN: All right. Endorsement. Yes/no.

MS. IBRAGIMOVA: Overall suitability

for endorsement. Does the measure meet NQF criteria for endorsement? One, yes. Two, no.

(Voting.)

MS. IBRAGIMOVA: Results are 96 percent yes. Four percent no.

CO-CHAIR THRAEN: Good. All right. Moving on to the next one is -- hold on.

MS. KEANE: 0567.

CO-CHAIR THRAEN: Yeah. Thank you.

MS. KEANE: It's Nicole Keane from Abt Associates, one of the measure developers.

CO-CHAIR THRAEN: Hold on. Excuse me. Hold on. There's been some reordering. Hold on a minute.

(Comments off record.)

CO-CHAIR THRAEN: 0097. We have National Committee on Quality. Their second one. I apologize. There it is. Medication -- 0097, Medication Reconciliation Post-Discharge.

NCQA, yes.

DR. GIOVANNETTI: Hello again.

So, this measure is a measure that

looks at the reconciliation between a hospital discharge medication list and an outpatient medication list occurring within 30 days post-discharge for adults 18 and older.

This is actually we have preemptively taken the move to combine two measures into a single measure here.

So, we have a provider-level measure of medication reconciliation post-discharge that's reported through the PQRS program, and a health plan-level measure of exactly the same that's reported in the HEDIS set for Medicare advantage plans.

So, this measure is not necessarily based off of a systematic evidence review, but really that best practice that medication reconciliation post-discharge is a critical component of transitional care.

Older adults are discharged -- older adults, and younger adults as well, are discharged from the hospital often with multiple medication lists that can be conflicting with

their outpatient medication list. And it is critical that the outpatient provider understand what medications were they prescribed from the inpatient setting and then reconcile that with the outpatient list to determine what's the appropriate list the patient should be on.

So, the measure uses medical record review once again to identify whether or not the two medication lists were reconciled within 30 days of discharge.

You'll notice that there are two different testing forms for this because this is tested at both levels. So, the provider level where we have done inter-rater reliability between different abstractors looking at the records and then at the health plan level where we've done construct validity analysis where we've compared the measure with other measures of medication management and found strong correlations, as well as reliability of the overall score of the precision of the estimate and found that the measure is highly precise.

As I said, it's used in different programs depending on if you're using the provider level or the health plan level.

So, for providers it's used for the PQRS program. For health plans it's used for Medicare advantage reporting.

The data that you have in here actually for the health plan-level reporting comes from special needs plans reporting, because that is up until -- actually, have they approved --

DR. BARTON: Yes.

DR. GIOVANNETTI: Okay. Good. I can say this. Up until a few days ago this measure was only approved for special needs plans reporting. But as of a few days ago, it has been approved for reporting by all Medicare advantage plans, not just the special needs plan.

And reported for all age groups in Medicare advantage 18 and older so that we are capturing that 18 to 64 dual-eligible population in the Medicare.

So, those were two significant improvements to the measure that we just got approved.

CO-CHAIR THRAEN: Chris.

DR. COOK: Yeah, this is Chris Cook. And going through this from the evidence standpoint as you pointed out, there is not any overall systematic review from what's there, but all the studies do consistently point towards the benefits of performing medication reconciliation particularly for patients who are at that risk standpoint of transferring between facilities.

Studies have all been primarily linked to medication reconciliation to reducing medication errors. However, no studies have actually linked medication reconciliation to morbidity or mortality simply because you have a number of steps that are within that.

And the developer states that most of those studies are basically just underpowered to get to that level.

As a personal note, when I look at

this overall being a pharmacist, when we look at what we spend in the national from a healthcare we're looking at 2.7 trillion dollars.

Drugs have consistently been approximately ten percent of that. So, that's 270 billion dollars that we spend on medications each year.

Back in 2009 NEHI actually did a report around medication non-adherence or medication misadventures. And their estimation at that point was 290 billion dollars.

So, in essence, as a society and as a healthcare system, we are spending basically one dollar fixing drug errors for every dollar that we spend on medications.

So, if you look at it from a purely economic standpoint or an economist would look at it, the question would be why in the world do we even pay for any drugs whatsoever?

As we are all practitioners in this room, I think if we took away that tool in our toolkit to improve patient care it would be

considered ridiculous.

So, to me, the next thing that we have to do is we have to put this as a national priority to pay better attention to the medications that we use, minimize the adverse effects that are there and be proactive in pursuing patient care in this area.

CO-CHAIR THRAEN: Lisa.

MS. MCGIFFERT: Thanks. What we're measuring here is actually that a check was done, but is there any indication that when a check was done there was actually reconciliation?

DR. GIOVANNETTI: So, this is something we actually in our past reevaluation of the measure looked into in depth of how can we get at the quality of the medication reconciliation, which includes more than just, you know, were they reconciled. Did you do education with the patient about it? Did you actually identify that each medication was indicated?

As we looked through all of our

options for actually getting down into this level of detail, it was just not feasible with a quality measure. And this may be one of the areas where the limits of measurement are actually you can't get up with this.

So, one area that we did actually we didn't look at this, but we looked at some research of others where they compared the hospital discharge medication list to the list in the outpatient record to then say, okay, well, what was the number of discrepancies?

The challenge with that is that you can't tell if the discrepancy was intentional or not. So, someone might have actually -- a provider in their outpatient list might have discontinued something that someone was discharged with because they didn't feel it was appropriate. And that was an intentional discrepancy versus an unintentional discrepancy.

And so, getting down into that level of detail what's documented right now in medical records we really can't get to that level of

detail.

And so, where we are right now is with a measure that does just say, did you look at both lists and compare them?

It, I agree, doesn't get at the level of quality we want to, but I will point out that in special needs plans this low level is only being done in 35 percent, on average, of discharges.

So, we see a need to significantly improve performance on this measure and then look at other things potentially like structural measures and quality improvement efforts that can be used to help bolster this whole concept of medication reconciliation.

MS. MCGIFFERT: Thank you.

CO-CHAIR THRAEN: Ed, and then Josh, and then Yanling, and then Steve.

CO-CHAIR SEPTIMUS: It looks to me as reading this that this could be reconciled by any number of three professionals; is that correct?

I'm just going to tell you

philosophically something. It says that any one of three practitioners can do this.

It seems to me by doing it this way, that you've taken the physician off the hook for his or her primary responsibility for patients. So, I have a concern about that.

And then secondly, you've said within 30 days; is that correct? Is there a particular reason you picked 30 days? That assumes that a patient discharged will see a practitioner in 30 days, which of course is not always the case.

DR. GIOVANNETTI: So, I'll first address the three different practitioner types. This can be done by a physician, any type of prescribing practitioner, including a nurse practitioner, a clinical pharmacist or a registered nurse.

And this is based off of several interventions that have shown that these different professions can actually do medication reconciliation very effectively as part of team-based care. And so, actually there's a lot of

evidence that clinical pharmacists can play a very critical role in medication reconciliation.

What we look for is that it's documented in the outpatient record so that it can't be that a clinical pharmacist does a reconciliation and does it over here and it never gets documented back to the central record.

And so, that's kind of where we are looking for -- we don't want to say this has to be the physician that does this, but it has to be documented in the record.

CO-CHAIR SEPTIMUS: I'm not disagreeing with the value of other practitioners, but it seems to me that somewhere in the medical record that the physician has to sign off that medications have been reconciled. That's what I'm saying.

I know that pharmacists play a valuable role. I know that nurse practitioners and nurses, I mean, physician assistants can play a valuable role, but that's generally under the supervision of a physician.

And so, my only concern about this is not that they don't play a valuable role, but that I think we have to have someplace in there that that is acknowledged by the physician. That's my only thought.

CO-CHAIR THRAEN: So, I want to remind people that we're talking about the evidence right now. So, if your questions are related to the evidence --

CO-CHAIR SEPTIMUS: Based on the evidence, it sounds like you have the evidence supporting --

(Comments off record.)

CO-CHAIR THRAEN: Josh.

DR. RISING: Hi. Thanks. This is Josh Rising. I just wanted to make sure I was understanding the numerator and the denominator correctly on this.

So, if you have a patient discharged from the hospital who never follows up with his or her primary care physician, so -- then so med rec is never done. So, presumably that would

show up for both the health plan and for integrated delivery system as no med rec being done. And then would it also show up for the physician who has been assigned to that?

So, for the physician it's only when you see a patient who has been discharged from the hospital, did you do med rec at that particular visit?

DR. GIOVANNETTI: Yes, you're entirely right. There's different accountability models.

DR. RISING: Now, is there a reason we wouldn't want to be assigning it to a primary care physician assuming that the patient has a primary care physician as well?

DR. GIOVANNETTI: I think you just hit the nail on the head as identifying who is the accountable physician for a patient that never followed up with a physician.

And so, that's just a limitation of the PQRS program.

CO-CHAIR THRAEN: Yanling.

DR. YU: Yes. I didn't hear Ed's

question about 30 days. I do have a concern about that, too, because I understand that sometime you document it not right at the time or maybe have a little short window, but how do you distinguish between someone delayed their documentation versus, you know, not just the reason that, you know, got 30 days.

To me, I don't see any -- in most cases you don't need 30 days to -- because, you know, to verify the medical records, adequate patients and, you know, their caretakers.

DR. GIOVANNETTI: So, you're correct. The evidence for 30 days versus 15 versus seven, you know, there's not enough evidence to really say this is the one threshold.

30 days is the -- a lenient time. And as I said before, we still see performance at 35 percent for just even reaching medication reconciliation within 30 days.

So, when we've talked about this with our advisory panels and they've suggested shortening the follow-up time, they've actually

said maybe we should see performance improve before we make this a more difficult measure.

I'll also point out that two of the administrative codes that can be used to meet this measure are transitional care visits that were recently approved by CMS, which can occur within 30 days of discharge.

And so, part of this is aligned with the billing that CMS has said transitional care occurs within the 30 days post-discharge.

And so, those visits -- so, we're in some ways trying to align to minimize burden in that way as well, but it's definitely on the future when we see performance get to a certain level that we would want to raise the bar and makes this something that looks at a more stringent time level.

DR. YU: Thank you.

CO-CHAIR THRAEN: Steve.

DR. LAWLESS: The evidence that -- or the numerator/denominator excludes observation patients, or includes observation patients?

X percent of patients who are there within -- or seen out of the hospital within 48 hours for observation do not categorize as inpatient. They are not part of the denominator?

DR. GIOVANNETTI: Give me one second here as I'm thinking through. So, this is a discharge from an inpatient facility.

DR. LAWLESS: Uh-huh.

DR. GIOVANNETTI: So, if they are considered -- if they are observation and considered hospital outpatient, they are, you are correct, not included in this measure.

DR. LAWLESS: Okay.

DR. GIOVANNETTI: We have been definitely going through the efforts of how do you distinguish hospital outpatient claims that are observation stays that are actual stays versus hospital outpatient that is not a stay. And that's one of the challenges we've been struggling to overcome.

DR. LAWLESS: Okay. So, you're looking at it.

And the other question is, is one of the exclusions are patients who are readmitted before the reconciliation is done.

I would consider was the evidence -- that the reconciliation was not done caused the admission -- the readmission.

DR. GIOVANNETTI: I think that's a very good point. So, I will say that if they're readmitted, they get picked up in the measure the next time they're discharged. So, they don't eliminate from the measure whatsoever.

You are correct the readmission could be the result of the medication reconciliation, but we can't necessarily determine if they did have an appointment scheduled where this was going to happen and they came back to the hospital before that appointment could happen if it was a readmission within 24 hours, within 48 hours.

And so, we do exclude those readmissions assuming they'll be captured the next time they're discharged from the hospital.

CO-CHAIR THRAEN: All right. Any other -- should we vote on the evidence?

MS. IBRAGIMOVA: Importance to measure and report, 1(a), evidence structure process intermediate outcome. The votes are one, high, only eligible if QQC submitted; two, moderate; three, low; four, insufficient evidence.

(Voting.)

CO-CHAIR SEPTIMUS: Lunch is here. So, some of you have already -- so, I think if you want to peel off and just grab your lunch and come back to eat, that would be great.

MS. IBRAGIMOVA: The results are 36 percent high. 55 percent moderate. Nine percent low. And zero percent insufficient evidence.

CO-CHAIR THRAEN: All right. Performance gap. Chris.

DR. COOK: Yes. As already mentioned, there is definitely a performance gap especially with the special needs plan beneficiaries.

It is interesting to see that they're going to be expanding that to Medicare advantage

plans and having it all the way across the board.

And of course Medicare advantage plans do take into the totality of the patient care both from the medical and the pharmacy side. So, I think I see that as a good thing.

The special need plan beneficiaries were an average of 36.6 percent. The tenth percentile reported 9.4 percent. And the 90th percentile reported only 62 percent showing clear room for improvement both from the mean, as well as the variance across what's going on.

Within the PQRS system, interesting to me here is that only 1.6 percent of eligible providers chose to report on this measure. And, again, this is one of those to where physicians are able to choose which measures. It's not mandatory for them to choose on it.

And so, what you see there is that those who did report obviously have systems in place that allow them to do this in a very structured fashion.

And so, you saw a mean performance

there of 96.3 percent. So, I think it clearly shows this can be done. It's just not being done by most. And so, we have a lot of room for improvement.

CO-CHAIR THRAEN: Go ahead, Yanling.

DR. YU: Thank you. That's just one of my question. The statements that only 1.6 percent of eligible provider choose to report. So, I guess this performance gap is really how to make them to report.

So, I wonder if you have any thoughts on that how to make people to do things that you make measures on.

DR. BARTON: I'm glad to say that from what I understand of the MACRA legislation, there will be -- CMS is looking to hold physicians accountable either through what they've called APMs, alternative payment mechanisms, which means like an ACO or some sort of actual risk arrangement, or if not that, then through what they call MIPs, the merit-based incentive program, which is going to ratchet up over time

the requirements for clinicians in practice, wherever they're in practice, to organize themselves for reporting on quality and for quality improvement.

So, I think that there's a lot of reasons to be very hopeful. Of course, the distance between legislation that's passed and the regulations that get written can sometimes be as we're suffering in the Supreme Court even today so that it's not a slam dunk, but I think that there's reasons to be hopeful.

And also reasons if you're so inclined, to be active with professionals organizations to try and influence how this gets written in regulation.

DR. YU: Thank you.

CO-CHAIR THRAEN: Jason, then Susan, then Charlotte.

DR. ADELMAN: I'm not sure that I believe the performance gap. And that is to say that I'm not sure that I believe in the validity of the measure.

And so, they're intertwined, my comments, but I do believe that I think it was somewhere like around 35 percent of the time people are checking the box.

I just don't believe that 30 percent of the time when people are discharged from the hospitals, doctors consider the medicines that they're on.

I think that what's more likely is that they just don't care about checking the box. There's not enough meaning and, you know, it's egregiously bad care to, like, not consider the medicines that a patient that was just in the hospital was on. And I don't think that 60 percent of our providers are egregiously that bad. I just think they don't care about checking the box.

And so, you can call that a validity comment, which I can circle back around and make it again for a performance gap, but I just don't -- I don't believe it.

CO-CHAIR THRAEN: Susan.

DR. MOFFATT-BRUCE: I think just from the perspective of physicians -- so, as a CQO for a large academic medical center, PQRS is not very -- not been in our wheelhouse at all. PQRS measures have been chosen by administrators that find them easy to report on the behalf of the physicians.

As MIPs is rolled out, I am encouraged because I think the docs are going to have to pay attention and they are going to be made accountable, because money talks. And I think that's the first incentive for physicians, unfortunately, to get kind of on board.

I think the more PQRS measures that are meaningful, that are put in place and endorsed will be helpful, because we can choose the ones that will impact patient care and ultimately improve, you know, the systems approach to care.

So, I think that is encouraging. I think that is important. I think the MIPs program with the right PQRS measures are going to

be very influential in providing the care we want for our patients.

MS. MCGIFFERT: Can you tell me what that acronym stands for?

DR. MOFFATT-BRUCE: I'm just trying to remember. What is it? Merit-based incentive program.

And so, I think it's encouraging because they took away this idea that they were going to reduce all the Medicare reimbursement, but rather put in pay for performance and accountability metrics at the physician level.

Right now it's very much at the institutional level so they think it's my job, whereas now it's going to be down to the provider. It's going to be their job, which is great.

CO-CHAIR THRAEN: Charlotte, and then we have someone on the line that has a comment.

DR. ALEXANDER: So, I have a concern that this is not meaningful, because it's too easy to check a box and not do the activity, do

not really reconcile.

And even more important is to find out whether the patient filled the prescription, is taking the medicine, is taking it appropriately. And that needs to be looked at for disparities.

I don't know that this will be applicable for disparities, but I think it misses what we need to do.

CO-CHAIR THRAEN: And online is it Kimberly or Ann? It's Kimberly. Kimberly, you have your comment about PQRS?

DR. APPLEGATE: I already made it.

CO-CHAIR THRAEN: Sorry. Thank you.

Any other conversation about performance gap?

(No comments.)

CO-CHAIR THRAEN: Okay. Should we vote?

MS. IBRAGIMOVA: Importance to measure and report, 1(b), performance gap. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: The results are 36 percent high. 36 percent moderate. 23 percent low. Five percent insufficient.

CO-CHAIR THRAEN: All right. Reliability.

DR. COOK: The measure developer has done testing at the level of the measure score. It was performed both, as I said, at the physician level. So, that was done through charts. And then also from electronic medical level testing for the plan level. So, two samples were done for data element reliability.

In one sample, it was from four practices using 62 patients. And then in the PQRS system, 38,000 plus patients were used to examine.

The nominator rate of agreement was 96.8 percent, which indicates that the abstractors almost came to the same conclusion.

And then the numerator had a kappa score of 0.97. Obviously very high.

CO-CHAIR THRAEN: All right. Any questions?

Yes, Yanling.

DR. YU: Yes. Thank you.

I have concerns about how the documentation would consider as the fact that medication reconciliation had been done.

I think on Page 21 it says any five -- what it said that the data on which you perform any of the following evidence need criteria.

So, you document a list of five things, but my concern is that things missing is very important for medication reconciliation is to document any there's contraindication, have had a drug reaction, have had a, you know, communication with the patients and caretakers, you know, about, you know, it's a process to me rather than just a single document said I done, check the box.

So, the whole thing is how this process went through that makes sure that everyone on the team, including patient

caretaker, is on the team and understand this thing.

So, I'm concerned about that just single check box.

DR. GIOVANNETTI: I think that this panel has raised one of the central challenges with measuring medication reconciliation, which is this balance between we want to see that many things were done.

Like you mentioned, all of this process was done. As Charlotte mentioned, you know, all of the things happened to help people actually understand their medications.

What's actually documented, though, as -- sorry, I don't see your name down there, but what's happening -- Jason. What's happening versus what's document is -- there's a disconnect there.

And so, we try to strike a middle ground where we look for these types of documentation that is reasonable with what we expect a provider to actually write in the note

section of I looked at the, you know, here's the discharge list, here's the things that were changed, or here's a copy of the discharge list or, you know, or, as you say, checking the box or just a notation of reviewed and reconciled.

Those types of things are what we look for as kind of the bear minimum that we would expect of providers to document, but they may be doing much more that they're not documenting.

And we want to strike that balance between not asking providers to spend all of their time documenting 15 steps of I educated the patient, I, you know, I looked for contraindications, I did this, I did this, I did this, because that turns into 15 boxes we have to check versus also trying to have some type of minimum documentation.

DR YU: I just think that, you know, to ask for providers to check whether they have been, you know, indicated in the medical records in term of a cancellation that, I mean, a contraindication, a drug potential interaction of

and adequate patient, I think, is a huge investment actually in the time spent, because that can happen.

That can have a medical error and have, you know, that can cause more actually to fix the medical error. So, I think that that shouldn't be a, you know, a burden, look at it as a burden.

You can just make boxes, make prioritize what needs to be done and check the box. Thank you.

CO-CHAIR THRAEN: Chris, then Missy.

DR. COOK: Yeah, I think, Yanling, what you're bringing up is absolutely true. And the pharmacy profession, I know, is definitely working on this.

I'm involved with the Pharmacy Quality Alliance, where we are trying to answer and to create measures and to build more robust measures.

This current measure right now is sort of that basic first step that is where we are

currently.

There are a number of CMMI grants that are going on around the nation, which definitely deal with medication use and trying to get from where we currently are to medication optimization.

And so that work is currently being done, but we don't have the solutions yet to bear. And a lot of it comes into what are the data limitations, both from EHR as well as claims data, a number of different things that are happening that are the barriers.

And I think overall our healthcare system is trying to build that infrastructure to get past that bridge, but as of right now, this is the best we can do.

CO-CHAIR THRAEN: Go ahead, Missy.

MS. DANFORTH: Given, I think, the Committee has really warranted concerns about doctors just checking the box and the fact that the measure has been endorsed since 2007, have there been any efforts to just do some random

observations in some of these practices that are voluntarily reporting on this measure to actually see if the physicians or the providers, I guess, are just checking the box or actually going through the steps, and is that a consideration?

DR. GIOVANNETTI: So, we do not actually run the PQRS program, so we don't know which providers are reporting on this measure. That's a CMS effort.

Where we do know that there is effort is around the special needs plans. And particularly, because this is a HEDIS measure reported by special needs plans, and we've worked with CMS to actually better understand their quality improvement efforts around medication reconciliation and understanding what are some of the things that they are doing to try to improve this process overall, particularly as it leads to better outcomes for patients in the end.

So, I have not personally done any observations. It's not a study that we have done, but it is something that CMS is actively

interested in and is understanding what are the efforts going on in special needs plans to do medication reconciliation.

MS. DANFORTH: Wait, but I think my question is slightly different in terms of the reliability of the measure.

And so, this is giving physicians and/or health plans credit every time a box is checked.

And so, like, to Jason's earlier point, if they're just checking the box and not doing anything, that does speak to the reliability of the measure.

DR. GIOVANNETTI: So, just to be clear, this is not --- there's no box. This is actually looking for notation, as we said here, of they signed off saying, I reconciled this.

And we have not actually done an observational study of the reliability. I don't know, Mary, if you want to build on that.

DR. BARTON: Well, I'll just say the health plan measure requires chart review. And

the special needs plans are called upon, if audited, to document what words notated in a chart are supporting their supposition that this is a numerator hit.

So, I am less likely to be concerned about this use in a health plan setting than I am in a physician setting. And I can appreciate the wide frustration with the PQRS program, but I guess I would just echo what Susan said earlier that there's hope that things are going to change, and that this is a good time to see change.

CO-CHAIR THRAEN: Charlotte, did you have something, or is that left over? Okay. Let's vote. Reliability.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2(a), reliability. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: The results are: zero percent high; 68 percent moderate; 27 percent

low; 5 percent insufficient.

CO-CHAIR THRAEN: All right. Validity.

DR. COOK: Within validity, there was testing done as well. A systematic assessment of face validity was done very similar to some of our other measures in this area, where there was an AMA convened PCPI standardized process using 33 members.

The mean rating for face validity was 4.0, with 73.91 percent of respondents either agreeing or strongly agreeing that the measure can accurately distinguish good and poor quality.

When you look at the threats to the validity, there are no exclusions to this measure, and it is not risk adjusted.

When you do look, there is a 33 percent gap in the performance between the 25th and the 75th percentile. So, you do see a distinguishable difference.

And there was definitely a gap among the low and the higher performing Medicare

special needs plans when looking at it from that standpoint.

CO-CHAIR THRAEN: Jason.

DR. ADELMAN: So, they're primarily relying on face validity that had --- and please correct me if I have this wrong --- that had some 30 experts that were asked the question, do they think the measure would be valid? And the majority said yes.

However, I think, you know, depending on how you couch the question now that that has happened, they have some data. And as I said before, 35 percent are compliant.

I bet if you went back to the same group and said, do you believe that 35 percent of our doctors are not doing med rec, that they would all now find it not valid.

And so, I think that, in this case, face validity is not acceptable and perhaps investing in doing a more thorough validation is worthwhile.

And also to the point --- I'm sorry.

You didn't know my name, and now I don't know your name.

DR. GIOVANNETTI: Erin.

DR. ADELMAN: Erin. To Erin's point, you know, it would be one thing if the measure was about documenting of med rec.

And, in that case, if the numerator didn't claim it's actually measuring med rec, but it's measuring the documentation of med rec, then we can say, okay, we're at 30 percent and people have to, you know, improve their documentation.

But since it's claiming it's an actual measure of med rec, I just don't believe it's valid, and I don't believe the face validity test of 30 experts at that time without this information is good enough.

CO-CHAIR THRAEN: All right. Other thoughts?

CO-CHAIR SEPTIMUS: Erin, I want you to meet Jason.

DR. GIOVANNETTI: I just want to put out one thing, which is that it's not based

solely on face validity.

We did do a test of empiric validity where we looked at construct validity and its correlation with another measure of medication management, medication review, and we see strong correlations there. So, we actually are seeing this behave the way we would expect it to behave.

I would also say that this measure just --- this was the original panel that did the face validity assessment, but we just went through this whole process with our panel again because this went through reevaluation, and they continue to support the face validity of it.

So, I agree with your assessment, and they echoed the same concerns that there is a problem with documentation of medication reconciliation; however, I don't necessarily think that is unique to this measure.

I think that is a problem with all quality measures that are based off of chart review when you are getting at something that is not routinely documented, similar to what we had

around falls.

You may be asking about falls, but not documenting asked about falls, no falls documented.

I think it is something that is a challenge with all quality measures; they are really no measures of what's documented.

And we want documentation to improve, because that in and of itself is an important component of patient safety.

CO-CHAIR THRAEN: Steve.

DR. LAWLESS: Real quick. In your validation studies, did everybody agree to what medication reconciliation actually was?

DR. GIOVANNETTI: They agreed to what we could measure, which is what is documented here.

CO-CHAIR THRAEN: Should we vote?

MS. ARDIZZONE: I'm sorry. There was one comment on here that we got pre-reading this from the --- what is it? I'm so sorry --- oh, from the Nursing Home and Assisted Living

Consortium, saying that they should be removed from the denominator because it -- I guess it wasn't generalizable to their kind of group.

Could you comment on that, or have you seen that comment?

DR. GIOVANNETTI: So, this is the same comment that was made on the falls measure; it's the same organization.

When providers are providing care in the -- a non-acute inpatient setting like a custodial nursing home, they may be eligible to report this measure if it is selected for them to be reported on.

And I think that they recognize that it is a challenge, it's a burden, but I don't think they provided very good clinical evidence that because someone is discharged to a nursing home they are not --- they shouldn't have medication reconciliation done.

CO-CHAIR THRAEN: All right. Let's vote for validity.

MS. IBRAGIMOVA: Scientific

acceptability of measure properties, 2(b), validity. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: The results are: 0 percent high; 59 percent moderate; 41 percent low; 0 percent insufficient.

CO-CHAIR THRAEN: All right. We're in the gray. Moving forward. Feasibility.

DR. COOK: There are no concerns within feasibility. This is not an eMeasure, but the data is captured from electronic clinical data that is being used to report for CMS meaningful use program.

Let's see. At the health plan and physician level, it's obtained through the administrative claims, electronic clinical claims for patient --- or paper medical records. So, no concerns overall.

CO-CHAIR THRAEN: All right. Any questions?

(No questions.)

CO-CHAIR THRAEN: Shall we vote?

MS. IBRAGIMOVA: Feasibility. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: And the results are: 32 percent high; 59 percent moderate; 9 percent low; 0 percent insufficient.

CO-CHAIR THRAEN: All right. Usability. Chris, do you have any comments?

DR. COOK: Just what's already been stated in the fact of its NCMS Medical Part C special needs plans. And apparently we've just been told it's been extended to all of Part C Medicare advantage plans.

Also, it's within the NCQA ACO accreditation program, and that may also be expanded to be used in the State of the Healthcare report, as well as the Quality Compass reports for public reporting.

CO-CHAIR THRAEN: Yanling.

DR. YU: Yeah, I just have one

comment, a question. The measure was first endorsed in 2007 and is then recent endorsement is 2012.

And since 2007 and, you know, with the time progress, have you seen any efforts being taken or any progress in transparency accountability once you started to see endorsement?

DR. GIOVANNETTI: So, in the special needs plans, we've definitely been seeing an improvement in performance, particularly when the payment has been attached to this as part of the CMS Stars program. That has definitely led to an improvement in performance on this measure, which I think we demonstrated.

I think it's been about a 10 percent -- just over the last year, it was a 10 percent improvement in performance.

In the PQRS program, I think we've talked at length about the challenges there where we don't see the variation to see the improvement, but we hope to see it as the program

itself improves.

DR. YU: Thank you.

CO-CHAIR THRAEN: All right. Shall we vote?

MS. IBRAGIMOVA: Usability and use, the votes are: one, high; two, moderate; three, low; four, insufficient information.

(Voting.)

CO-CHAIR THRAEN: We're missing two. Try again.

MR. ANDERSON: Kimberly, we need your vote.

DR. APPLEGATE: Hi. I'd like to vote high.

(Pause.)

MS. IBRAGIMOVA: The results are: 33 percent high; 48 percent moderate; 19 percent low; 0 percent insufficient information.

CO-CHAIR THRAEN: And the last one is suitability for endorsement. Yes or no.

MS. IBRAGIMOVA: Overall suitability for endorsement. Does the measure meet NQF

criteria for endorsement? One, yes; two, no.

(Voting.)

MS. IBRAGIMOVA: The results are: 60 percent yes; 40 percent no.

CO-CHAIR SEPTIMUS: Okay. Well, that one was a close one, wasn't it?

(Comments off record.)

CO-CHAIR SEPTIMUS: Okay. Skipping around based on people's needs and flights, we're going to go to 0537.

MS. KEANE: Hi. This is Nicole Keane with Abt Associates. We are the measure developer; CMS is the measure steward. And I have on the phone as well my colleague from Colorado, Dr. Gene Nuccio, and also Dr. Angela Richards.

Dr. Richards will start us off with an introduction.

DR. RICHARDS: Thank you.

CO-CHAIR SEPTIMUS: If you can make your introduction very brief, we'd appreciate it.

DR. RICHARDS: Yes, I can make it very

brief. Thank you.

Older people receiving home healthcare have relatively higher rates of falls. And those are, in turn, associated with injuries, healthcare resource use -- including ED use and hospitalization --- and increased mortality rates.

We talked in review of a previous measure about the American Geriatrics Society and the British Geriatrics Society clinical practice guidelines which make -- recommend use of a multifactorial fall risk assessment.

We also have evidence based on a Cochrane Review that found that risk assessments associated with reduced rate of falls from health providers are in a unique position to assess the environmental and other circumstances within the patient homes that may increase falls risk, and then to provide interventions and recommendations to mitigate those risk factors.

This process measure encourages use of a systematic multifactorial assessment for falls

risk and provides home health agencies and consumers with information that will enable them to monitor the extent to which fall risk assessment is conducted for ambulatory patients.

It is not limited to older adults; however, we should short of note that 82 percent of home health agency users are over 65. So, it does really hit on that population pretty well.

The measure is calculated based on data from the mandated OASIS-C data set that the home health agencies collect as part of their comprehensive patient assessments.

And then the definition of the measure is the percentage of home health episodes in which patients who can ambulate had a multifactorial fall risk assessment at start or resumption of care.

So, I think I'll just kind of conclude that there since we've already heard a lot of the evidence as it supports other falls risk measures.

CO-CHAIR SEPTIMUS: Thank you. So,

who's going to do this one?

CO-CHAIR THRAEN: Pat Quigley gets to do this.

CO-CHAIR SEPTIMUS: Pat does it? All right. So, let's go through the evidence, Pat.

DR. QUIGLEY: Sure. Thank you. And I'd also like to thank the developers so much for that brief overview.

And what I would like to say, in relationship to the evidence, is that this is a home healthcare indicator.

And the evidence that is presented is essentially for those who are community-dwelling adults not in home care.

So, putting that aside, I will say, though, that I did go ahead and add to the literature review to support this. Because -- because this is a home health indicator, that there is evidence to address falls beyond those who are ambulatory especially, you know, in looking at those who are not ambulatory, but there's different risk assessment that has

advanced since 2012 when this was endorsed that looks at intrinsic and extrinsic risk factors inside the home and outside the home.

So, even though the evidence that's here is good in relationship to why it's important to identify fall risk in older adults, I think, you know, the evidence in home care is really emerging.

So, that being said, lack of evidence is not evidence that something is lacking. So, it's still an important indicator because there's a lot of opportunity there, but this is a very important measure for home care.

And I hope my CMS colleagues, if you want to comment on my comments, that's okay.

CO-CHAIR SEPTIMUS: This is an endorsed measure, correct?

DR. QUIGLEY: Yes.

MS. KEANE: Yes, it is.

CO-CHAIR SEPTIMUS: So, are you trying to tell me we have an endorsed measure on no evidence?

DR. QUIGLEY: No, it is, but it's based on AGS guidelines, which is for ambulatory care people in the community, but this is a home care measure.

CO-CHAIR SEPTIMUS: Right.

DR. QUIGLEY: Right --- which is an extension into the home setting.

CO-CHAIR SEPTIMUS: Any comments on the evidence?

(No comments.)

CO-CHAIR SEPTIMUS: Okay. Well, I guess -- well, we got to wait for Laura to come back. Take your time, Laura.

MS. IBRAGIMOVA: So, importance to measure and report, 1(a), evidence structure process and intermediate outcomes. The votes are one, high, only eligible QQC submitted; two, moderate; three, low; four, insufficient evidence.

(Voting.)

MS. IBRAGIMOVA: The results are: 10 percent high; 70 percent moderate; 10 percent

low; 10 percent insufficient evidence.

CO-CHAIR SEPTIMUS: Okay. We will move on. Gaps in care.

DR. QUIGLEY: Yes. In terms of the opportunity for improvement, the data that was submitted on this did not indicate much opportunity for improvement.

The data for the year ending in June of 2014 for the agencies indicated they had 20 valid episodes in which there was 98.4 percent performance, but recognizing that this measure is that for those who can ambulate in home care that they have a multifactorial assessment done. So, it's really yes or no.

DR. NUCCIO: This is Gene Nuccio from the University of Colorado. Just to clarify, we have close to six million episodes of care annually from 2011 through 2013 calendar years. So, we have more than 20 episodes on which these data are based.

And the --- for the --- it's a three-part --- it's --- we have --- while the measure

specifically looks at the assessment piece, we also have data elements that look at whether or not that information was used in the plan of care, and whether or not it was used --- if it was implemented -- that is, that actions were taken.

And those data show that all across those three years approximately between 96 and 98 percent of agencies are doing the assessment, the plan and the doing of this.

DR. QUIGLEY: Thank you.

MS. ARDIZZONE: So, 98 percent of the people are already doing this?

DR. NUCCIO: --- of agencies are, in fact, assessing the fall risk.

CO-CHAIR SEPTIMUS: Yes, go ahead.

DR. NUCCIO: Just to give you some context, we have two historical --- one historical piece of data for the patients. Approximately between 28 and 30 percent of patients come to home health with a history of falls, defined as two or more falls, or a serious

fall in the last 12-month period.

After the fall risk assessment, approximately 88 percent of patients in home health are judged to be at risk for falls.

So, the assessment process is identifying patients who indeed are at risk using that multifactorial falls risk assessment. And agencies are taking action, again, at a rate of about 98 percent to put it in the plan of care and to do something about it.

As a result of this, the actual percentage of patients who go to the hospital for emergency care due to a fall, a serious --- obviously a serious fall -- is only about 7 percent of those patients who go to emergency departments for any reason.

So, the entire assess, plan and do process that we have in place seems to be very effective in reducing that potential risk population that is 88 percent, down to about 7 percent who actually experience a fall that requires an emergency department help.

CO-CHAIR SEPTIMUS: Lisa.

MS. MCGIFFERT: So, is this a measure that's been topped out and --- it sure sounds like it. And what do we do with that?

CO-CHAIR SEPTIMUS: Well, I think the first thing we ought to do is vote on the gap. And then if people think it's topped out, then we'll discuss next steps. Does that make sense?

So, I don't see any other hands up, so let us now vote on the gap.

MS. IBRAGIMOVA: Importance to measure and report, 1(b), performance gap. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: And the results are: 5 percent high; 25 percent moderate; 60 percent low; 10 percent insufficient.

CO-CHAIR SEPTIMUS: It didn't pass. Okay.

DR. PINES: Right. So, at this point we would decide whether or not to put it on

reserve status or, I guess, do we go through all the other measures?

Okay. So, we go through all the other ones first, and then we can decide at the end.

(Pause.)

CO-CHAIR SEPTIMUS: If it didn't meet 1(a) and 1(b), then that was an automatic stop. So, okay. Let's keep going then.

DR. QUIGLEY: For reliability, because this is electronic clinical data is the source of this data. So, for reliability, as mentioned by the additional comments, it was electronic clinical data that was used for this testing, from July 3rd, 2013 to June of 2014, at the facility agency level. And this is where there was the 9,443 agencies that tested 3.8 million patients.

And they had the critical elements, and they did the reliability testing interclass correlation of 0.91. So, there's good reliability.

CO-CHAIR SEPTIMUS: Comment. Lisa,

it's your --- okay.

(No comments.)

CO-CHAIR SEPTIMUS: Seeing none, let's vote on reliability.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2(a), reliability. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: The results are: 41 percent high; 50 percent moderate; 9 percent low; 0 percent insufficient.

CO-CHAIR SEPTIMUS: Okay. Now, we go to validity.

DR. QUIGLEY: Thank you. There's no issues with validity. It's a yes or no indicator of whether or not assessment was done. So, there's no issues with validity.

CO-CHAIR SEPTIMUS: Comments.

(No comments.)

CO-CHAIR SEPTIMUS: Seeing none, we will vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2(b), validity. The votes are one, high; two, moderate; three, low; four, insufficient.

(Voting.)

MS. IBRAGIMOVA: The results are: 27 percent high; 64 percent moderate; 9 percent low; 0 percent insufficient.

CO-CHAIR SEPTIMUS: Feasibility.

DR. QUIGLEY: The data is easily collected. It's collected in the clinical electronic data elements, as I had mentioned. And it's collected through the clinical registry, the nursing home MDS, the home health OASIS program. So, it's feasible to get this, and it's not an eMeasure.

CO-CHAIR SEPTIMUS: Comments?

MS. ARDIZZONE: Sorry. I had a question for the developer. I had brought this up at another measure. So, again, there are some competing measures. There's 0035, 0101. I know each of the developer's responses has been,

"Well, I'm looking at a very specific or a very different aspect, or literally a different piece."

A patient goes through all pieces as they're going through their healthcare. How do we -- again, I'm concerned about the burden on people reporting, and we're measuring all different aspects of one fall, or one person. So, is there any talk, especially as we need to make sure that measures follow a patient through all transitions of care?

DR. RICHARD: Yes, it's a very good point. We have been using this particular one now since 2009, and it has been in the chart to remind home care staff particularly to do their risk factor assessment, and just focus on that.

This is not -- this is a process measure. This isn't an impact measure. There is help and work on developing a measure that crosses care settings for an outcome measure. And so, that may be where some of the harmonization comes in.

So, it is a very good point. At this point, this is a measure that the home care agency population is very comfortable with, very familiar with, and is -- you know, whatever question you're answering from a provider perspective, the burden doesn't really fall on the patient so much as it does the provider, and this is the item that the providers have been collecting for many years now.

DR. NUCCIO: Also, the fact that this is the community, if you will, measure. That is that the assessment is done in the patient's home, which is far different than the other settings that the patient comes from, like a hospital or nursing home. And so, there are many other factors that need to be assessed in a unique way for -- by our nurses when they do in and do this assessment as part of the comprehensive assessment in home care.

The data has mentioned part of the OASIS instrument, and have been on the instrument since 2010. So, we're very comfortable with

assessing it, but looking forward to harmonization across post-acute care settings with the impact measures.

CO-CHAIR SEPTIMUS: Iona?

CO-CHAIR THRAEN: So, we have to enforce that. So, these have a three-year cycle. I would anticipate in three years that the impact work should be close to being done, and that all of these various measures that are either all related to falls or all related to nutrition, or all related to whatever came across the sectors, the impact measures should be compared, and we should see some retirement going on as we go to a common measure that might have multiple attributes.

So, part of that problem is because of the information systems that are different across each sectors. Also, the information systems are going to have to change in order to be able to yield that new type of measure. So, I think in three years, we will be at that place. At least starting.

CO-CHAIR SEPTIMUS: Okay, I don't see any other hands. So, we'll vote on feasibility.

MS. IBRAGIMOVA: Feasibility: The votes are one high, two moderate, three low, four insufficient. The results are 48 percent high, 52 percent moderate; 0 percent low, 0 percent insufficient.

CO-CHAIR SEPTIMUS: Okay, now, usability?

DR. QUIGLEY: Thank you. I think that everybody has discussed -- described adequately that this is being done.

So, how it is being used to now move into quality of practice I think is the next step, and that was also my comments in getting to present this is that there is emerging evidence in the home health arena that the falls that happen inside the home are different than outside the home, and maybe they'll use this as an opportunity to do something different and build upon a yes/no measure, and also look at the emerging fall risk assessment tools for the home

population.

The Missouri Alliance of Home Care has done extensive work in this area. So, I think there's an opportunity to really build upon this clinical setting where patients are vulnerable, and not just those who can ambulate.

CO-CHAIR SEPTIMUS: Thank you, Pat.

DR. YU: Thank you. I have a question. Currently the measure is used for healthcare public reporting. What if this measure was not approved; would the data continue to be collected? Would there continue to be public reporting?

DR. QUIGLEY: Well, it's in OASIS. So, it is something that's being done anyway. I don't know about public reporting, but it's an OASIS element, because OASIS has been requiring this, as mentioned, since 2010. This measure was originally adopted in 2009 before the OASIS.

DR. YU: So, my question is it has been -- I've been thinking about the re-endorsement. What is new? When you look at the

time change and what has been changed, do you compare the improvement now days and compare it with data from back in 2008 when you first endorsed?

So, I just -- I guess my question is what is new in this endorsement that when you continue public reporting? Will the public say there's something different or improved since 2007?

DR. QUIGLEY: I'm not the developer, but I'm say what's new is that they are doing it. People are being assessed for falls; those who are ambulatory. Back in 2007, it wasn't. Now they are being assessed, and maybe the developers would want to comment as well.

DR. NUCCIO: Yes, yes. We are. There's always been a very strong level of assessment, but that has actually improved slightly across the three years of data. Three complete calendar years of data that we have.

Also, I would like to point out that -- you raised the question about whether or not

the outcome measure would be -- well, in this case it's a process measure. Would be posted on Home Health Compare.

CMS has the option, as I understand it, to post measures on Home Health Compare that are not NQF endorsed, and one of those was the dyspnea measure that was not endorsed by NQF, but continues to be reported on Home Health Compare because we believe in the value of that particular outcome measure.

But I'm a contractor. I can't speak directly for CMS.

CO-CHAIR SEPTIMUS: Go ahead.

DR. RICHARD: This is Angela Richard, one of the measure developers. I'd like to point out that one of the big differences in this measure is that it requires a multi factorial, and standardized validated assessment instrument.

So, I think since 2007, not only are they doing it, but they're doing it in a more systematic way and using evidence-based tools actually and recommendations.

So, it really has -- we've seen not only an improvement in what people are doing, but in how they're doing it we believe.

DR. NUCCIO: Including the Missouri Tool that you referenced.

DR. RICHARD: Right.

DR. NUCCIO: We know that there was a major push by another private commercial contractor to create or find a multi factorial instrument that was validated and standardized, and they settled on and have highly touted the Missouri document.

We know that many agencies make use of that. CMS does not require a specific tool. Only that it is standardized and elevated.

DR. QUIGLEY: Thank you. I did add that. This is Pat Quigley's voice. I did add that to the comments in the review.

CO-CHAIR SEPTIMUS: I think this committee knows that I think people prefer NQF endorsement, but it is not a requirement. So, I think it's going to be up to the decision of CMS

if we reserve this because it may have topped out whether or not they want to continue to have reporting. So, let's go to usability.

MS. IBRAGIMOVA: Usability: The votes are one high, two moderate, three low, four insufficient information. So, the results are 14 percent high; 67 percent moderate, 19 percent low, zero percent insufficient information.

CO-CHAIR SEPTIMUS: Okay, now we'll go to the last one, which is should we overall suitability for endorsement?

MS. IBRAGIMOVA: Overall suitability for endorsement: Does the measure meet NQF criteria for endorsement? One yes, two no.

CO-CHAIR SEPTIMUS: Just vote yes or no, and then go to the next step. We have to go through the whole process. If you decide to endorse it, we'll go back and say whether we want to put it on reserve status, okay? Just to let you know, okay?

DR. RISING: My understanding is that it can't be endorsed because it didn't qualify

for that -- it didn't meet --

CO-CHAIR SEPTIMUS: That's -- that --

DR. RISING: We still have to go through the other steps though because it's still eligible.

CO-CHAIR SEPTIMUS: That's how you should go to the rest. It didn't pass the gap. Why don't you introduce yourself since you've joined.

MS. MUNTHALI: Hi. My name is Elisa Munthali. I'm Vice President for Quality Measurement at NQF. I'm stepping in for Helen.

We do need to apologize about the process you went through for reserve status yesterday. It was a misstep. What you should do, the friction you're sensing and the vote reflected that there wasn't an opportunity for improvement there.

It sounds like from how the votes are falling that you still think it is a good measure. So, what we're going to do is vote on the overall suitability, which you have just

done, and it has passed. Now, we're going to go to a yes or no on reserve status.

CO-CHAIR SEPTIMUS: That's just what I said. Of course, it's not what I thought yesterday.

MS. IBRAGIMOVA: So, the results are 67 percent yes; 33 percent no.

CO-CHAIR SEPTIMUS: Passes the consensus. So, now we have to go to whether or not we want to retire this to reserve status. Do you have -- oh, you've got it. Look at this. Not yet. Wait a minute. Okay, so now we get to say do you want to endorse the maintenance potential to reserve status.

The votes are one yes; two no. The results are 95 percent yes; 5 percent no.

CO-CHAIR SEPTIMUS: All right, it's on reserve.

CO-CHAIR THRAEN: So, because of scheduling issues, we're not going to go back and re-visit the old one yet. We might end up doing that on the phone after the fact.

We're going to go to 419 because we have the developers on the phone, who have to leave at what time? Developers have to leave at 1:30, and it is now 1:30. So, are they leaving?

Okay, so which one is it? It is 419. They have to leave at 2:00. 419, please step up. This is Quality Insights of Pennsylvania. Good afternoon. Please introduce yourselves.

DR. BERG: Is the microphone working? Well, I was prepared to say good morning, but good afternoon to the group here. I'm Sven Berg. I'm the Chief Medical Officer at the West Virginia Medical Institute, which is the parent organization for Quality Insights.

I am joined by Alyssa Crawford from Mathematica, with whom we work now. On behalf of CMS, and as the measures developer for this measure, Quality Insights of Pennsylvania is pleased to introduce NQF 0419: Documentation of Current Medications in the Medical Record for Consideration of NQF re-endorsement.

This measure was developed to promote

medication safety by requiring physicians to review patients' most current medications list at every encounter.

This allows for more effective monitoring for medication errors, and is a critical activity to prevent adverse events.

The measure received initial endorsement from NQF in 2008. It was implemented into the Physician Quality Reporting System, beginning in 2010, and into the Meaningful Use Program, beginning in 2013.

In 2013, over 100,000 eligible providers who participated in the PQRS program reported to this outpatient measure using either claims or registry data.

The intent of this process measure is for all eligible medical professionals to document a list of current medications, using all immediate resources available at every encounter or patient visit.

This list must include all known prescriptions, over-the-counters, herbals and

vitamin/mineral dietary nutritional supplements, and must contain the medication's dosage, frequency and route of administration.

This measure focuses on the adult population, those 18 years and older, and the denominator includes all visits occurring during the 12 months reporting period.

Patients in an urgent or emergent medical situation, in which time is of the essence and the delay in treatment would jeopardize their health are excluded or exempt from this measure.

Evidence suggests that frequently identifying document DNA, maintaining a list of patients' medications requires -- reflects high quality care. The process this measure assesses is foundational to multi component approaches to decrease ADE's, which in turn reduces unnecessary medical treatment, and minimizes the financial burden on patients and payers.

Eligible professionals see this as an important measure and many have opted to report

it in the PQRS and Meaningful Use programs. Furthermore, recent testing has shown that the measure is feasible, valid and reliable.

Unfortunately, our colleagues at CMS may not have been able to join us for this discussion of this measure due to the change in time, which was considered. However, based on our recent conversations with them, I believe they would want us to reiterate that this measure has an important part of a comprehensive approach to quality improvement and a key component of the GPRO, PQRS, and Meaningful Use BP programs.

Thank you for this opportunity to discuss the measure today.

CO-CHAIR THRAEN: Kendall?

DR. WEBB: So, I had one general question. I didn't find -- in this, you talk about eligible professional as part of this. Who do you consider an eligible professional?

DR. BERG: Sure. The EP's are those defined by CMS under the Physician Quality Reporting Program. I can give you an entire

list, if you like. I can read it, but it's a quite extensive list and it is defined by CMS.

MS. CRAWFORD: I would just add that the list includes physicians, as well as a number of other types of practitioners, including PA's, nurse practitioners, social workers, dietitians, audiologists, and also therapists including physical, occupational and speech therapists.

DR. WEBB: How about medical assistants?

DR. GREEN: This is Dan Green from CMS. Medical assistants would not be included in this.

DR. WEBB: So, licensed practitioners, anybody who carries a license it sounds like?

DR. BERG: That would be correct.

DR. WEBB: I just wanted to make sure we knew what we were talking about and who this applies to. So, 419 --

CO-CHAIR THRAEN: Steve?

DR. LAWLESS: Certified medical assistants cannot do this, correct?

DR. GREEN: We're looking for people who are submitting billing claims, if you will, to CMS.

CO-CHAIR THRAEN: Go ahead, Kendall.

DR. WEBB: I had one more quick question. Is this an e-measure?

DR. BERG: Yes.

DR. WEBB: All right, so evidence. They provided a great introduction. I don't think we need any more introduction. So, for evidence, this is a process measure. There was a systematic review, and NQZ presented.

There is good evidence that adverse drug events are a problem, especially in the outpatient setting. Of note, this is an attestation process, where the practitioner attests that they have reviewed or provided this documentation.

It does not tie back, actually. They have seen an improvement in performance just from an evidence perspective, but it has not tied back to a decrease in adverse drug events, as they

have seen an increase in attestations. I just want to make that note as we go through the evidence.

I think otherwise, the evidence is excellent. It does tie med rec to ADE's, but I -- what they're trying to do is decrease the ADE's, and they don't have that tie back in there. So, as we're voting, I want that to be noted.

CO-CHAIR THRAEN: Steve, did you still have a question? Questions? Josh?

DR. RISING: Hi there. This is Josh Rising. Certainly, there is a lot of evidence about the frequency of adverse drug events that is documented.

I guess what I didn't see, and I was hoping you could talk me through a little bit, was a sense of whether this frequency of med rec will help address the ADE challenge. Because that was something that I didn't see in the evidence, and presumably is an essential part of why to have the measure.

DR. BERG: Let me begin by making a slight modification to your statement, but this is not a medication reconciliation measure. This is a documentation of medications.

So, it is -- this is not a med rec measure in and of itself. One of the things that is difficult about any analysis of this type of measure is that this is a quality best practice, and to design a study that would compare two groups: one doing this and one not doing this, we really feel would be an unethical proposition.

So, obtaining the type of evidence that you're looking for is difficult for us to do, and could be unethical.

However, at the same time, we believe that a -- that it is logical to draw a conclusion that it is necessary to have the complete list of medications if you're really going to assess whether there could be adverse drug reactions, and then to be able to avoid those reactions.

MS. CRAWFORD: A number of the studies cited in our evidence back up the statement that

a lot of current documentation of medications is inaccurate. I think that's pretty clearly noted.

Sorry about that. I was saying in connection with the point that Dr. Berg just made, there's some pretty clear evidence that current medication lists are inaccurate, and that there's often incomplete or inaccurate information, not only about which medications a patient is taking, but the dose of the medication and administration, and other very important details.

I think Dr. Berg said it well in terms of the inability to devise a randomized controlled trial that would actually determine the effectiveness of this particular practice. I wanted to add that in addition, it is difficult, if not impossible, to design an analysis that could control for all of the other variables that might be inherent in the process, even if you wanted to look at whether there was a correlation with decreased and adverse events, it is very difficult to be able to control for all of the

other factors that influence that kind of an outcome.

So, it's the reason why our evidence and testing approach did not consider looking at that kind of a predictability assessment because of the fact that it would be very difficult to devise one that would accomplish that effectively.

MS. MCGIFFERT: So, just to follow up on that, is there evidence presented that doing this action leads to accurate lists as opposed to inaccurate lists that we have lots of evidence there is inaccurate listing?

DR. BERG: I'm not aware that we have anything to that effect.

MS. CRAWFORD: I would add that I think part of the reason why we don't have anything to that effect is because nobody is looking for that evidence because it is a sort of understood part of the process.

I was saying I think part of the reason why we weren't able to find evidence that

clearly made that statement is because of the unethical issue that Dr. Berg pointed out, which is that if this is such an established part of care that should be occurring in all visits, but which our test results show is not, that it would be very difficult to devise.

Nobody is going to do a study to show that connection because you would have to stop doing it for a certain number of patients.

MS. MCGIFFERT: You could have a third party verify the accuracy of the list versus inaccuracy of the list. There wouldn't be anything unethical about that.

MS. CRAWFORD: True. There are a number of studies that have looked at adverse medication reconciliation. Again, this is a medication documentation measure, but after medication reconciliation, there are studies that show that doing that process improves care, and that looking at the medication and discussing medication lists with patients does identify a number of discrepancies.

CO-CHAIR THRAEN: Lisa?

MS. DANFORTH: I just want to make one more comment, and I don't know if the other committee members who were here last year would agree with me, but I remember some radiology measures where the measure was basically, "Do you document this thing, or do this thing?"

What we agreed was that exposure to radiology is extremely important. It is something that should be measured and we need to measure. But this particular one was really too distant from the outcome.

This measure kind of reminds me of that, where medication errors are extremely prevalent, and we need good measures of medication errors. We don't have a good outcome measure. So, we're seeing these processes' structural measures.

But to me, the distance from the outcome is too far. In the case of those radiology measures, we said, "We're going to say no today, and we hope that encourages you to come

back with something better."

I'm just thinking that in my opinion that is what I'd like to see the committee do today. If think if we endorse this as a measure but we don't push this any further, three years from now this is going to be up for reendorsement and we're not going to be any further along.

We know medication errors are frequent and dangerous, and is this really the best way to improve on it? I just wanted to mention that. To be consistent also with the way we've treated some of our measures, I think.

DR. GREEN: This is Dan Green from CMS. Completely respect what you just said, but I would suggest that in the absence of a perfect or better measure, not to necessarily discount this measure.

I mean as a physician, it kind of shocked me as we were developing this measure that we even needed a measure to tell providers that they should be assessing what medication, herbals and botanicals that their patients are on

when they're caring for them.

As an OB/GYN, you'd be surprised how many folks, for example, are taking botanicals that could interfere with their clotting, but their docs have absolutely no idea about, or even the primary care doctor thinks, "Oh, you are on the same June 26, 2015 medications."

But in the mean time they see someone different, and they put them on something different. We talked about different types of practitioners. I mean even taking it, for example, to a dietitian who is trying to help somebody, perhaps, with weight loss.

It would be helpful for them to know if the patient is on an SSRI medication. They may be fighting a little bit of an uphill battle. So, again, it surprises me that we need the measure, but we do need the measure. And until something better is created, I would suggest this is important.

CO-CHAIR THRAEN: This is Iona. So, I guess the argument we just made talks about

individual practitioners, who are using supposedly CPOE kinds of tools and digital prescribing, and a variety of data sets that are out there.

On behalf of that, their individual expertise, they are documenting that in their individual EHR's. So, the argument that you made is that that should be available to the next provider, which is an interoperability question, and a transfer of information question, and not a documentation question, in my mind, of the provider that is doing the service at that point in time.

So, does this really get at that argument that you just made about the dietitian being able to see that you're on an SSRI, when it's really an interoperability question as opposed to a documentation question?

DR. GREEN: Again, you make an excellent point, and certainly at CMS and HHS in general as I'm sure you do, we all strongly support the use of electronic health records.

And equally or more important, the interoperability of the records.

Unfortunately, as you painfully know probably, we're not there with interoperability. I mean, things are improving. Pharmacies are starting to be able to communicate to electronic health records. Not just the pharmacy where the patient fills medication A, but if they go to a different pharmacy to fill another medication.

With the payers and what have you becoming, we'll eventually get to the point where that information will be fed back to the rendering docs system. But unfortunately, between that and the interoperability, we are just not there yet.

DR. BERG: I think to tag onto that, I would argue that the reason we're asking all eligible providers to do this is for that very reason: that the interoperability isn't there.

So, we're asking all of them to ask these questions, and to document that, and that makes that connection where interoperability

isn't there.

DR. GREEN: One last quick thing. As I'm sure you all know, the level of update of EHR use is different among different types of providers. Psychiatrists, for example, are some of the later adopters, the numbers aren't as high for them, for example.

Dietitians, if they're working in a hospital: Obviously, they're likely to be on EHR. But if they're in independent consultant? Not necessarily because many per say doesn't apply to them.

MS. CRAWFORD: This is Alyssa Crawford again from Mathematica. I just wanted to address a point that was raised about developing measures that get farther than this current measure. I wanted to just state that I think it is a measure that has been very clearly heard, and something that -- perhaps I should note that this measure is being maintained under a contract from CMS to develop and maintain electronic quality clinical measures for eligible professionals.

So, we have currently in the pipeline a number of measures that are under development. Some of which are very highly related to patient safety and go beyond the documentation of medications.

I would add that in order for those measures to be able to work, we need to make sure that the data that's going into those measures is accurate. Right now, the fact that a number of those providers are still not updating the medication lists at every visit suggests that those measures may be based on bad data, which calls into question not only their validity but their feasibility moving forward.

So, I think in many ways, this measure is foundational to that measure and what we're currently seeing in the performance doesn't suggest that we're at a point where this measure is ready to go away. We still need to make sure that this data is going to be accurate so we can measure the important things that you've brought up. They're really what we should be focusing on

moving forward.

CO-CHAIR THRAEN: Chris then Charlotte then Michelle.

DR. COOK: I think all the points that have been brought up, and this comes back to the previous measure when we were talking about medications is that we don't have great measures currently.

I think it hits on that face validity that there's no way to accurately find out what the patients are if you're not asking. Much like Dr. Dan Green was saying, it is shocking that people aren't doing this as an automatic part of their process.

It is not that you're seeing one physician. It is that most of our patients are seeing a multitude of physicians. So, whether you're the primary care who is looking at it and not getting the information back from their endocrinologist or their cardiologist or the podiatrist, if you're one of the specialists who is not on the primary care, then you still have

to be able to catch all the different points that are ongoing.

So, to me, it is almost as hard to believe that somebody would put in a central line without using aseptic technique, and I cannot believe that we have to put that as a measure, but it is something that is absolutely necessary now to help us avert further adverse drug events that we know are occurring and we have great documentation that is out there and is prevalent.

DR. LAWLESS: I'm going to give you the other side. This is just as much of a patient as a provider issue. So, before we just get on physicians on saying, "How dare you not enter the data." Evidence is out there. Articles published. Half the patients, more than half the patients, don't even know what medications they are on.

If you ask a family member, "What are you taking?" They are getting it wrong over half the time. Unless you ask them to bring the medicines in with them. You still get to about a

90 percent rate of accuracy.

So, I think what you're asking is you're trying to get a central repository of medications that we know you are taking from pharmacy data and from other stuff.

So, I think it is a great goal. But I also think quite honestly sometimes with medication listing the less data you have is better because it helps people question at the time of validity. "What medications are you actually taking? I have very minimal here."

So, the reality is the surgeries and that kind of stuff. That's why they ask at the last minute. It's because you can't trust the list. Even if you're asking, half the time the family doesn't even know what they're all taking.

Is it better to prompt the question at the time of delivery? Is it better to have a central repository that people can go to do this? Before we get on physicians only, it is just as much a patient and provider issue.

DR. GREEN: Well, there's no doubt

that that's true. I don't think that should let the provider off from making the effort to try to find out. We've even added language in the measure last time we went before NQF. I think it was something to the providers' best efforts or -- I mean if somebody forgets to tell the doc that they're on a particular medicine, they are not going to go revoke their Medicare or billing privileges or something like that.

I mean, everything in a medical record, as you will know, is based on the history that you're given. So, you can only do what you can do. But you're right; it definitely works both ways, but at least we can adjust the doctor part.

CO-CHAIR THRAEN: Charlotte and then Michelle.

DR. ALEXANDER: So, I'm hearing what everyone is saying, and I agree to a great extent. What I find is that none of us are perfect historians, and patients are not perfect historians.

They will come in one day and tell me one thing, and then on another day another. I get great benefit when I go to the operating room and my anesthesiologist has asked what drugs they're on and what allergies they have and I can compare it to what I have, and it's not always the same.

So, I think every time we ask is a benefit. It gives us a chance to catch things that may fall through the loops. It may not be perfect in an ideal world. But we've got transparency, so we can look.

My heparin office calls the pharmacy if the patient can't tell what they're on. We're really trying hard to capture this data. I know I would say every time we can it's a benefit.

DR. SCHREIBER: Thank you. I fully agree with you that we need a medication list and it should be in there every time. I'm trying to clarify about the documentation piece, and what you're actually measuring. Because according to what you've written, it's the licensed

professional who has to document the med list. Well, in our clinic the MA documents the med list, or a nurse, or a pharmacy technician.

Does that count when they do that and the physician reviews it? Do they actually have to do the documentation?

DR. GREEN: No. That's a great question. It doesn't really matter who does the documentation. Your medical assistant can do it, as you said. We do expect for the doctor to report this, but they would review it. "Okay, so I see you're on, you know, whatever." But they don't have to literally do the writing or the typing in the health record.

CO-CHAIR THRAEN: All right, I think it is time to vote.

MS. IBRAGIMOVA: Importance to measure and report 1A evidence, structured intermediate outcome. Votes are 1 high, only eligible if 2 QQC's submitted; 2 moderate, 3 low, 4 insufficient evidence.

Results are 19 percent high; 57

percent moderate; 14 percent low; 10 percent insufficient evidence.

CO-CHAIR THRAEN: All right, performance gap?

DR. WEATHERS: Performance gap: They show interesting evidence that actually using this attestation or they got better performance, they increased from 75 percent using PQRS and other measures but PQRS being the main one, from 75 percent attestation rate in 2010 to 88 percent in 2013. Eighty-eight percent is not 100 percent. So, I would say that there is a performance gap.

CO-CHAIR THRAEN: Any questions? Let's vote.

MS. IBRAGIMOVA: The importance to measure and report 1B performance gap: The votes are 1 high; 2 moderate; 3 low, 4 insufficient.

The results are 43 percent high, 33 percent moderate, 19 percent low, 5 percent insufficient.

CO-CHAIR THRAEN: The reliability?

DR. WEBB: So, for reliability, they actually have -- using data from Part B Medicare claims, PQRS, administering data from registries and EHR reports: Used 3 outpatient physician practices from 255 patients in the report there for their testing.

MS. CRAWFORD: The 255 patients was the number that was manually abstracted for the validity comparison and recorded later for the reliability. It was based on the full sample, which included 40 providers with an average of 407 patients and 770 encounters per provider. So, it's a fairly substantial size --

DR. WEBB: Okay, so, given that size when they calculated out the reliability, they had a reliability between 0.97 and 1, which is good reliability.

CO-CHAIR THRAEN: Lisa?

MS. MCGIFFERT: Just quickly. The data elements that we're looking at here are that somebody said they checked it? Right? Okay.

DR. YU: I have a comment. One of the

criteria for physical -- oh, this is reliability. I'm sorry. Never mind.

CO-CHAIR THRAEN: All right, we'll vote.

MS. IBRAGIMOVA: Scientific measurability of properties to reliability: The votes are 1 high; 2 moderate; 3 low; 4 insufficient.

The results are 50 percent high; 52 percent moderate; 14 percent low; 0 percent insufficient.

CO-CHAIR THRAEN: Sorry. Trying to multi-task. Didn't work.

DR. WEBB: So, validity testing. The specifications do align with the evidence. The validity testing was done both on the element level and the score level.

Again, we talked about it. Three outpatient offices and 255 patients were manually abstracted. It was noted in the discussion with the previous committee meeting that as far as the validity was concerned, face validity results at

the performance level were not reported and there was no risk adjustment.

There was no power analysis for the reported sample size either.

CO-CHAIR THRAEN: Any questions? Comments?

MS. CRAWFORD: I just wanted to point your attention to Question 2B 2.2, which indicates that we determined via simulation that our sample of 255 cases had a greater than 80 percent power to detect at least substantial kappa scores between EHR extract and --

CO-CHAIR THRAEN: Any questions? All right, vote.

MS. IBRAGIMOVA: Scientific acceptability of measure properties 2B validity: The votes are 1 high; 2 moderate; 3 low; 4 insufficient.

Results are 10 percent high; 71 percent moderate; 19 percent low; 0 percent insufficient.

CO-CHAIR THRAEN: All right, next?

DR. WEBB: Feasibility: the have four years of PQRS reporting. I don't see any issues with feasibility.

CO-CHAIR THRAEN: Go ahead.

MS. ARDIZZONE: I just had a question since this is four first e-performance measure. When do we -- or e measure. Do we just look at the technical review that you provided with the comments, and that's it? Okay, thanks.

DR. YU: Under feasibility, the criteria is "Can be implemented for performance measure?" My concern is that we have this discussion about how to really make this list as accurate. I mean documenting is an excellent idea, and you have to be accurate in order for the list to be useful.

I can see either way. My personal experience is I went to my mom's doctor, and the medication she was not taking anymore was still on there. So, it takes me going through with her with the doctor and hasn't been there for some time.

So, I'm just concerned about when you try to do performance measurements, you don't have a way -- or maybe you have a way and could explain it. How do you really make sure that list is accurate, updated and -- you know, so it's useful?

DR. BERG: You make an excellent point. Before coming to my current position, I worked in the DoD system, and we dealt significantly with medication reconciliation and the electronic medical record that we had, and it was very difficult. Medication lists were often wrong.

So, it really took repeated encounters with the patients, going through the list of everything that we have, striking those things that were on the list and adding new drugs, etcetera.

So, in this specific measure, there is no mechanism built into it to ensure that everything is correct. What I would argue is that if we don't ask at each encounter, we're

less likely to actually find those items that we need to find as we go along.

MS. CRAWFORD: I would just like to add to that that there's a lot of promising practices for how to improve the accuracy of medication lists, and we've talked a little about some of them: encouraging patients to bring their medications with them to their visit.

I think the reality is there's a lot of different ways of doing that, and rather than being prescriptive in this measure and directing providers to one type of workflow for how to accomplish it, we want to make sure that they're documenting it at every visit.

I think there's a lot of other work going on to ensure that they're using the right processes, which may not be at the level yet where you can call them best practices, but that help to encourage the collection of the right types of data at that point of care.

I think I agree with the points Dr. Berg raised. It is a very valid question, and I

think this measure won't fully accomplish the point that you're trying to make but I think it gets it into the workflow in a way that encourages providers to be thinking about how to maximize that interaction with their patient.

CO-CHAIR THRAEN: Jason, you had something?

MR. GOLDWATER: Sure. So, since this is an e measure, and I know one of the things I said this morning was to look at how feasible it is to do in the daily workflow of care; what I would say is these are a lot of excellent points that have been raised here, but if you have really looked at electronic health records over the past several years, the functionality that is probably the most robust at the moment is electronic subscribing.

It has been for some time, largely because it relies on a singular code set and a singular method of transport: getting information from one system to another.

It is remarkable that we're doing that

with medications, and we can't seem to do that with anything else.

CO-CHAIR SEPTIMUS: Jason, it is also required.

MR. GOLDWATER: Right, right.

CO-CHAIR SEPTIMUS: That's one of the three elements in it: Accountability and how it drives change.

MR. GOLDWATER: That is correct. I'm talking from a purely technical standpoint because I'm essentially referred to at NQF as a gearhead. Take that for however you'd like, as have I, usually by Helen.

So, I think the functionality is robust enough that even though we're not at the point where there's 100 percent accuracy on the list because we're still waiting for full bi-directionality between a patient and multiple prescribers based on where care is delivered, and that needs transmitted back to the primary care physician.

We are, as it was stated earlier by my

colleague at CMS, moving closer to that idea. And we will be there probably before we will be there with anything else at the moment.

So, even though there is the question of is there a way to guarantee 100 percent accuracy, which I think is probably an unreasonable goal in any element of this, I think we're much closer to that electronically than we will be on anything else, and it also goes back to even though this measure almost seems as if, "Why do we need something that almost seems so obvious?" At this moment, we definitely need a measure like that, and the functionality will catch up to deliver exactly what you want, probably sooner than it will on any other level.

CO-CHAIR THRAEN: All right, any other comments? Lisa is thinking about it.

MS. MCGIFFERT: Well, I do want to -- I mean I think I've heard about these kinds of measures for at least five years. Maybe eight years. I hear what you're saying about, "It's coming around the bend."

But I kind of feel like as long as we keep endorsing these measures that really don't tell us what we need to know, we're not going to be developing the measures that we really need to have that tell us what we want to know.

So, I see it as sort of a delay tactic. I understand what you're saying, but I just feel like this doesn't move us forward. It just is telling us that -- it's telling us a conversation took place maybe but it doesn't give us any sense of whether or not it helps a patient, improves care, avoids adverse events, is accurate. All of those things I'm not seeing here. So, that's my biggest concern.

CO-CHAIR THRAEN: Anymore comments? Shall we vote?

MS. IBRAGIMOVA: Feasibility: The votes are 1 high; 2 moderate; 3 low; 4 insufficient. Results are 15 percent high; 75 percent moderate; 10 percent low; 0 percent insufficient.

CO-CHAIR THRAEN: Usability?

DR. WEBB: Use and usability: Again, we've covered most of this. It is currently in use. Has been for four years. We've seen improvement. It is also a meaningful use criterion. So, it's used in meaningful use. A little bit different in meaningful use, but used just the same.

It is publically reported. There is information on improvement over time that we've already discussed, and so as far as unintended consequences, the only unintended consequence I can see from the way that this is worded is that anybody in the office can document these meds, and nowhere is it stated that the physician actually sees it before they write the prescription, and they write at least one prescription per visit based on the evidence.

So, you know, are we documenting but more having the most important person review it? It's not covered in here. That would be by only concern.

CO-CHAIR THRAEN: Any questions?

Let's vote.

MS. IBRAGIMOVA: Usability and use: The votes are 1 high; 2 moderate; 3 low; 4 insufficient information.

Results are 15 percent high; 55 percent moderate; 30 percent low; 0 percent insufficient information.

CO-CHAIR THRAEN: All right, and suitability for endorsement.

MS. IBRAGIMOVA: Suitability for endorsement: Does the measure meet NQF criteria for NQF endorsement? 1 yes; 2 no. Just missing one.

CO-CHAIR THRAEN: Missing one.

MS. IBRAGIMOVA: Results are 70 percent yes; 30 percent no.

CO-CHAIR SEPTIMUS: Okay, I'm going to turn this over to Jessie is just a second. I think I hear a theme about measures like this that cannot be directly tied to an action, but directly impacts patient care. Is that what other people are hearing?

I said there seems to be a theme emerging with some of these measures that it doesn't necessarily link to an action. I'm talking about an action by the physician -- I'm sorry.

The theme I'm hearing is that we're checking off a box, or we're doing something for the measure. We can't really guarantee it links to an action by the physician, which in turns links to better outcomes in patients. That seems to be a theme.

I don't want to take this as a criticism, but I think we're getting, and I think Jason has articulated this. We are getting to the point now where maybe we can start making some of those links.

Right now, we're left with measures that leave us a little bit unsatisfied for what really matters, and that's what's actionable and affects patient care. Other people think differently. It's whether we send that message. I don't know how we do that, but I think unless

other people -- it is one of the things that I hear coming out of this meeting.

DR. LAWLESS: This is also to Lisa's point, which emphasized the same thing. In some of these, you can't say no to them, because you feel like you're going against motherhood or something. At the same time, we're not really raising the bar, and saying, "Let's be leaders rather than laggers in this."

The technology can do this fairly easily, but is this really leading healthcare to a better place versus, "Okay, we're going to gradually get to this?" I think it's saying what you said. Lisa's point was well taken on that too.

DR. BURSTIN: And again, anything you guys state that is sort of conceptual, we will include in the report. I think the developers pretty clearly heard it is highly unlikely measures like this that point through the next time they come up for maintenance unless we've moved significantly forward, and also just to

plead all of you: If you know of innovative measures that are actually pushing the needle and kind of getting us closer to that place we want to be, please encourage people to bring them in.

Talk to us. We can work with those folks. Again, we just want to keep making sure there are better and better measures coming in. Some of these are great measures. They are kind of legacy measures in some ways, and I think there's a place for them.

We don't want to throw the baby out with the bath water, but we do want to continue to bring in the innovative ones that I think push the envelope more.

CO-CHAIR SEPTIMUS: Everybody agrees we will make that comment in our report, and maybe Helen or others can comment. NQF really has moved the bar into innovation to help innovate new measures, and you may want to comment on that relatively recent effort by NQF to do that.

DR. GREEN: Sorry. This is Dan Green

from CMS. I'm going to drop off but I just wanted to thank you all for your consideration of this measure and certainly suggestions to improve it we would be all ears. Thank you again. Good luck with the rest of your meeting.

DR. BURSTIN: Thanks, Dan.

CO-CHAIR THRAEN: So there's lots of efforts. We are trying to both be involved in the measurement side. Like, for example, the issue this morning we talked about antimicrobial stewardship.

We've got an action team that has launched with help from Ed and lots of support to really begin thinking about not just what the measures are but what are the evidence-based strategies, how do we pull together the remarkable NQF membership to be much more focused on action.

How do you take a measure and run with it. How do we really drive meaningful improvement. It's been pretty unsatisfying at times. To feel like, okay, here's the measure and

it's kind of thrown over the transom and we should wait to see things happen. I think there is definitely a sense we would like to be more engaged.

The standing committee can be engaged in all those efforts really helping us think not just about the measures but how they get used. Even the CDC measure this morning we actively reached out for, not sort of passively waiting. It's really important. It's a national priority where do we get those measures working with Ed and others to really go after it.

DR. BURSTIN: I want to make a comment about that, Helen. So for the maintenance measures that are scheduled that we know and can anticipate, I'm wondering if there has to be another set of criteria that makes an evaluation of these issues that we are identifying sort of up front to say, "Okay, this measure has been in play for six or eight years and it's still too far away from the goal," and whatever, whatever. But there is at least some sort of assessment of

that.

CO-CHAIR THRAEN: Yeah and actually, I believe we're going to launch it with the new project? Elisa is shaking her head. When our new projects begin in the fall, we hope, all of our CDP work will actually -- we'll actually have a different process for measure maintenance. We shall see.

We have found this sort of painful as you have to go laboriously through some of these older measures in great detail. In the future maintenance first of all will move from every three years to four years just to lighten that load a bit in terms of these maintenance measures.

The emphasis on it will be overwhelmingly around gap in care and usability and use. We will only re-evaluate evidence if it's changed and we will only really look at the testing and scientific reliability, validity, etc. We keep either change level of analysis, change data source, made it an e-measure.

We are really trying to streamline that because we feel like we have to, first of all, give us a break. This is all intense at times. But also really emphasize that point; measures coming back to us should really be about are you helping to move the needle on quality.

If you're not, why not, and it'll be trying to sort of move it to more feasible data sources and get much more emphasis from the field. Is this measure helping? Is it hurting? Just part of usability as well as evidence of unintended consequences and hopefully give a little more breathing room to try to bring in more newer innovative measures.

MS. DANFORTH: I had a quick request for that. So in doing that, there are some measures we looked at today where when the measures were first endorsed disparities were identified. When the measure was brought back to us for maintenance, the disparities are still noted so it didn't seem like there had been any work on the disparities in between.

Can that be part of the performance gap piece, that like it's expected that if the disparities are identified that within the three or four years something is done to address that. I think that is really important. I think it looks strange if the committee keeps re-endorsing measures for these noted disparities, like 12 years later they're still there.

CO-CHAIR SEPTIMUS: Well I think it gets to the concept is that when we see that we want to know that the measure is actually changing or moving the needle and that is something that we did not necessarily see. So I don't want to labor this but I just thought conceptually that what I'm hearing and it sounds like everybody seems to be really invested in out of this meeting and going forward.

Now, we have -- it depends upon when people are leaving and how long we have a quorum, how many measures. I'm going to let Jesse sort of walk you through some things.

Oh, I'm sorry. Yanling, did you have

--- I'm sorry.

DR. YU: I just want to make a quick comment. Since I'm new, when I'm looking at all the measures, there's a endorsed since 2007, for example, and then re-endorsed 2012. Then I start looking at what has been done, what is new, what accomplishments.

I really like to have your ideas where you have them talk about in the spreadsheets, really list what has been done, what is new, what accomplishments, and that would be the gap actually, what re-endorsement is needed.

DR. PINES: Thanks. So we have two more measures that are on the agenda. We have 45 minutes. We want to see if we are going to lose quorum before 3:00 so can you please raise your hand if you are going to have to walk out before 3:00 to get on a flight. How about 3:30? We are going to lose --

And for Ann and Kimberly?

DR. APPLEGATE: I'm here. I'm okay. It's still early here in California.

DR. PINES: Okay, great. So we will try to get through the last two measures if possible.

CO-CHAIR SEPTIMUS: Okay, Iona.

CO-CHAIR THRAEN: 2732. Is that correct? INR Monitoring for Individuals on Warfarin after Hospital Discharge. CMS and Mathematica are the developers. Who is the representative?

DR. RISING: I am for that one.

CO-CHAIR THRAEN: Josh is. Okay. Thank you. You want to introduce, please?

MS. CULLEN: Yes. Thank you again. This is the second measure developed by FMQAI under previous contract with CMS. As the current contractors Mathmatica is representing CMS' interest. We'll note any questions we are unable to answer at this time we work with CMS and FMQAI to get answers for you.

Measure under consideration, INR Monitoring for Individuals on Warfarin after Hospital Discharge. That's the percentage of

adult in-patient hospital discharges to home for which the patient was on warfarin and discharged with a non-therapeutic INR who had an INR test within 14 days of hospital discharge.

This process measure is a hybrid with data source from both an electronic health record and Medicare administrative claims. Use of the hybrid meets two goals; the use of novel techniques for measure reporting, as has been discussed here, and they reduce burden on providers.

Warfarin continues to be widely prescribed. It has a narrow therapeutic range and needs to be monitored closely to lower the risk of complications such as thromboembolism or bleeding. Current guidelines recommend follow-up for out-of-range INRs, particularly those less than two and greater than three.

This measure is focused on these at-risk patients out of the therapeutic range as they transition from the in-patient setting to the home from a controlled to a less-controlled

environment.

Public comment raised a question of attribution of post-discharge care to hospitals. This was discussed with this committee yesterday. Dr. Ahlen mentioned -- had talked about this so hospitals should be concerned and held responsible for this type of follow-up.

FMQAI undertook extensive field testing at seven geographically and characteristically diverse hospitals. The liability scores calculated at the hospital level provided an indication of the ability to distinguish between signal and lies and Mathematica augmented FMQAI's work in this area like calculating a measure level reliability score that indicated the ability to distinguish performance among hospitals.

Criteria and validity tests showed strong agreement between electronic and manual abstraction for all of the data elements required to be captured by the EHR. Although discharge status showed slightly lowered strength of

agreement.

Field testing identified issues related to the accuracy in identifying discharge medications in the EHR. As a result of this, the measure was modified to include a proxy to look for the administration of warfarin in the hospital on the day of or day prior to discharge.

Construct validity tests confirm the specification captured patients discharged from warfarin that nearly all patients should have had a follow-up INR. Feasibility tests demonstrated that all data elements were found to be available in the EHR systems and used by the hospitals which included Epic, Cerner, and McKesson products.

This measure is not yet in CMS programs, nor has it been reviewed by the MAP. Planned use is for CMS' hospital and patient quality reporting program. There are four measures that address warfarin and INR monitoring and this measure uniquely addresses the care transition and need for appropriate post-

discharge follow-up. Thank you for your consideration.

CO-CHAIR THRAEN: Josh.

DR. RISING: Hi. I think that is a very helpful summary. Should we just start right in on discussing the evidence? Sounds great. All right. We'll keep things moving forward.

I think the big question for me that I would like to hear the developers talk about a little bit is the selection of the INR range that is used here in particular. I think we all know that your target is two to three for most patients and the evidence presented kind of looks at -- well, there is evidence that looks at a number of different ranges at which adverse events occur.

A lot of studies kind of at five and above. Some look at anything outside the two to three range. So I would like to hear a little bit kind of on the selection of the 1.5 to four as the area to target the follow-up because I didn't see a lot of studies that did target that range

in particular as where to be focusing priorities.

DR. CROGHAN: Sure. So the short story is that's a synthesis of a lot of different studies that you mentioned. One cut of it is that for people with heart valves the therapeutic range is 3.5, so a little bit higher. You will recognize that this is now .5 above and below that broader range two to three and a half.

Having said that, a lot of the evidence is built upon sort of standards of your three to five, five and above, particularly for bleeding events. This was selected by the expert panel as sort of a conservative estimate of where that line ought to be but there is no clear standard.

But the lower end I think the standard is pretty clear. That's where the cuff links are for a real jump in the adverse outcome. The robotic events at the lower range. It's a little bit less clear at the upper range where there's a jump in the number of bleeding points.

DR. RISING: I'm curious to hear

anything you can share about why the panel didn't choose five because it does seem that there are a number of studies that really kind of use five and above as the -- where they look at the adverse events.

DR. CROGHAN: That may be a nuance I can't provide you since we weren't in the room.

CO-CHAIR THRAEN: Lynda.

DR. SMIRZ: I have a question for the developer when I was reading through this. One of those is I don't see a correlation with picking 14 days. Some of the evidence that you cited said that there was an increased rate in mortality at one year if it was 45 days between the time that the INR was drawn, not 14 days. So is there something that I missed where there was a correlation between 14 days and extra adverse events or is that when the adverse events occurred?

DR. CROGHAN: The 14-day standard was chosen based upon the American College of Chest Physician Guidelines where if you had a slightly

out-of-range INR that you should be tested -- retested within seven to 15 days. That was felt like it was giving providers a benefit of the doubt using the best available evidence.

DR. SMIRZ: In this measure here there was no evidence that was cited that showed that. Was that --- you're just basing that on --

DR. CROGHAN: No. That's right. It's based upon the guidelines.

DR. SMIRZ: Okay. Thank you.

CO-CHAIR THRAEN: Other questions about the evidence? All right, we'll vote.

MS. IBRAGIMOVA: Importance to measure and report on an evidence-structure process and to media outcome. The votes are one high, only eligible if QQC submitted; two moderate, three low, four insufficient evidence.

CO-CHAIR THRAEN: Vote again. Did we get 21?

MS. IBRAGIMOVA: Twenty-one.

CO-CHAIR THRAEN: Twenty-one. We're there.

MS. IBRAGIMOVA: The results are 33 percent high, 52 moderate, 14 percent low, zero percent insufficient evidence.

CO-CHAIR THRAEN: Performance gap.

DR. RISING: Okay. So the evidence that's presented by the measure developer show that when they tested this in seven hospitals, they found -- I'm sorry, a mean rate of around 50 percent for meeting the measure. It certainly shows that there is room for improvement in this particular measure.

CO-CHAIR THRAEN: Any questions?

Lynda, you still have your card up. Do you have a question? Okay.

Anybody else? All right. Let's vote.

MS. IBRAGIMOVA: Importance to measuring and report 1(b) performance gap, the votes are one high, two moderate, three low, four insufficient. The results are 33 percent high, 57 percent moderate, ten percent low, zero percent insufficient.

CO-CHAIR THRAEN: Feasibility.

Reliability. Sorry. It takes two.

DR. RISING: Great. So on the reliability testings, again, there were seven hospitals that were assessed and five of them have scores that were above the acceptable threshold when it came to reliability. Two of the seven that did have smaller sample sizes were below the specified threshold for reliability.

CO-CHAIR THRAEN: Any questions?

Missy.

MS. DANFORTH: So it looks like the reliability testing was dated from 2011 to 2012 used CMS data. I know that measure is not being like actively used in any programs, but is it actually being used by a health plan or state?

DR. CROGHAN: Not that I'm aware of.

MS. DANFORTH: So you just -- so because you had access to the CMS data, is that the reason we did seven hospitals?

DR. CROGHAN: Remember this is a hybrid measure so we had to know who was discharged on warfarin using EHR data. The CMS

data is to determine the numerator.

MS. DANFORTH: Okay.

CO-CHAIR THRAEN: Any other questions?

Pat.

DR. QUIGLEY: Thank you. I would just make a quick correction. On page nine for the numerator is says the patients are the denominator so if this goes anywhere, that would be the patients in the numerator.

CO-CHAIR THRAEN: Note to staff, or developer.

All right. Any other questions? Let's vote.

MS. IBRAGIMOVA: Scientific acceptability of measures not released to a reliability. The votes are one high, two moderate, three low, four insufficient. The results are 14 percent high, 71 percent moderate, 14 percent low, zero percent insufficient.

CO-CHAIR THRAEN: All right. Validity.

DR. RISING: Great. There were a

couple of interesting questions here. I would be curious to get the developers' thoughts here. The first had to do with discharge status so they went back to look to see because there were some important exclusion criteria including if you were going to a skilled nursing facility so they wanted to only look at individuals who were being discharged to go home or a few other places and it varied by hospital.

One hospital in particular there was only 70 percent accuracy between what was in the EHR and kind of where the patient was ultimately found to be going if I was reading correctly. I did have one other question but wanted to get the developers' thoughts on how that may challenge the measure.

DR. CROGHAN: My recollection is that where there was disagreement it was largely based upon chart reviews and that had to do with several where there was disagreement within the chart where the patient was actually discharged. For example, some patients were discharged to

home healthcare where other people were discharged home.

That was often based upon different recorders. For example, the social worker may discharge one person so it was sort of within that category. Overall I don't think that changed the overall assessment or the validity of the measure.

CO-CHAIR THRAEN: Lynda.

DR. SMIRZ: I had some concern for the developer that you might be able to answer. One of those goes back to exclusions as far as the threat to validity. A patient that was re-admitted within 14 days they may have been re-admitted because of a problem with not having their INR checked and they were on too much or too little Coumadin. Also patients that died within 14 days.

DR. CROGHAN: I can deal with the second one first. Very few people died probably does not have a significant impact, a measurable impact on the measure. We don't know the reasons

why people were re-admitted. That does make up about 25 percent of the exclusions.

CO-CHAIR THRAEN: Any other questions?

Josh.

DR. RISING: One other, I think, exclusion criteria to note is the readmission to the hospital. You look at how many patients were excluded because of that particular criteria, it was between 40 and 50 percent of patients were excluded because they were coming back from the hospital.

That may or may not be related to the fact their INR was so out of whack when they were discharged but it does seem like it's a lot of the population that we are trying to target with this measure who are being excluded.

By the time you add up, you know, people who are going to skilled nursing facilities or who have died or who have gone back to the hospital I think it's over half the patients who are being discharged with out-of-range INRs who are then excluded from the

measure. That did kind of raise a couple of flags for me.

MS. CULLEN: One of the reasons that this exclusion was included was because it was assumed that once the patients were readmitted they were going to get the INR so the onus of the follow-up was no longer on the hospital because of the readmission.

CO-CHAIR THRAEN: All right. Any other questions? Vote, please.

MS. IBRAGIMOVA: Scientific acceptability of measure that relates to validity. The votes are one high, two moderate, three low, four insufficient.

CO-CHAIR THRAEN: I need one more it looks like. Someone's got a lazy remote. Oh, she's gone. We've got one who stepped out. That's fine. Okay. Thank you.

MS. IBRAGIMOVA: The votes are 15 percent low, 60 percent moderate, 25 percent low, zero percent insufficient.

CO-CHAIR THRAEN: It takes a team.

All right, feasibility.

DR. RISING: The only things to speak to on feasibility are that, again, this is a measure that is drawn both from claims data and the EMR data but it seemed to be done successfully by the measure developers in this case not tried other questions on feasibility.

CO-CHAIR THRAEN: Questions? Vote.

MS. IBRAGIMOVA: Feasibility, the votes are one high, two moderate, three low, four insufficient. The results are 40 percent high, 55 percent moderate, five percent low, zero percent insufficient.

CO-CHAIR THRAEN: Usability.

DR. RISING: I do think the main question in my mind on usability has to do with because of the very small sample sizes that you're going to have from a lot of hospitals in part due to those exclusion criteria kind of how much -- how useful kind of it's going to be.

I know we discussed this a bit in the reliability section. We did have, you know, about

a quarter of the hospitals that had too few patients to feel like reliability was particularly good. So I do think there is going to be a challenge just on the sample size of this measure.

CO-CHAIR THRAEN: Ed and then Victoria.

CO-CHAIR SEPTIMUS: Just quickly I think we need to measure this. One of the unintended consequences of this measure is driving people to use other oral anticoagulants that do not require monitoring and whether or not that we have anything in place to look at.

DR. RISING: Ed, can you say that one more time and speak up a little bit?

CO-CHAIR SEPTIMUS: I'm sorry. I said one of my concerns is that even before this measure is that since you don't have to measure INRs that some of the new oral anticoagulants, are we going to drive increased use of oral anticoagulants and do we know what the relative comparisons are with those through the safety

methods.

DR. CROGHAN: I'm sure, Ed, that's a question you will be asking us for years and years.

CO-CHAIR THRAEN: Victoria.

DR. RICH: I have a question. I think it's related to usability. It might be the whole thing. In my 30 years of practice, primarily inner-city academic medical centers and a struggle with minorities, with INR clinics, and on and on and on. And so I thought it was interesting that you didn't find any difference there perhaps not looking at that.

But I think moving forward with the minority populations I think would be so key to see if we could stratify that or do something with it. I don't know if you have any comments. I think many people have seen that within the INR maintenance and trying to keep this going, even in the first 14 days.

DR. CROGHAN: Are you referring to the notion that there are some hospitals, and there

are some hints in the data that didn't point to statistical significance in regards to hints that there were some hospitals that had more troubles that had larger minority populations?

DR. RICH: I'm just looking at the usability to expand how we're using it.

DR. CROGHAN: Right. Sure. So, yeah, I think it does tell you that if you're not getting follow-up.

DR. RICH: Right. I just thought it might be an extra kind of add-on as time goes forward.

DR. CROGHAN: Hopefully that's a useful thing. Great.

CO-CHAIR THRAEN: Lynda.

DR. SMIRZ: Can I ask the developer just one final comment? I guess it falls under usability. I released a patient from the hospital and want them to have an INR. They were in town and they moved to a different place. What are your suggestions on how that -- what kind of a mechanism is the hospital going to put

in place to ensure that that happens? Any thoughts?

DR. CROGHAN: So as in any transitioning care what this measure tends to do is to have the hospital take proactive steps to improve outcomes before discharge. I think the term that we heard yesterday was the hospital should own that transition period as best they can.

If you have a patient who you know is moving and is not going to be able to get good medical follow-up, then the recommendation is that they stay as an in-patient until they do. You don't have to be a real -- they could go to the hotel across the street for that matter. But there are ways that you can help people get the appropriate follow-up so they are not out of range. Does that answer it?

DR. SMIRZ: Well, that is an answer. I don't know how satisfying an answer it is. I have the number of a company that has a number of hospitals along the border in Texas and they go

back to Mexico.

Then they come back when they have a problem because the care is better there. They come as our readmissions, etc., etc. But once they are across the river, we don't have that much control over them.

CO-CHAIR THRAEN: Kendall.

DR. WEBB: I agree with Lynda. I work in an inner-city -- as you guys know, an inner-city population. I think owning your own healthcare to some extent is what needs to be done here.

I have a lot of patients who I work in several of the hospitals in town and I'll see them at each hospital because they didn't get what they wanted at the last hospital and I'll be like, "Mr. Jones, I just saw you yesterday at this other hospital for this same problem."

And they are not following up. At what point do we not put necessarily all the onus on the hospitals but figure out a way to give the patients control back of their own healthcare in

some way, shape, or form.

CO-CHAIR THRAEN: Michelle.

DR. SCHREIBER: Thank you. I do agree. I work in an inner-city hospital as well and our patients travel but I have to admit I think that we do have the burden of responsibility if we discharge a patient with an INR out of range to find them. Maybe that's harder in some populations but I think that it is our responsibility.

CO-CHAIR THRAEN: Victoria, did you have another comment? Victoria, did you have another comment? Okay.

Anybody else? Vote, please.

MS. IBRAGIMOVA: Usability and use, the votes are one high, two moderate, three low, four insufficient information.

CO-CHAIR THRAEN: And finally, suitability for endorsement. Overall suitability for endorsement, does the measure criteria for endorsement, one yes, two no. The results are 90 percent yes, 10 percent no.

CO-CHAIR THRAEN: So, we're at the end of the hour, and we also have to have a public comment period so I don't think --

DR. PINES: So, I think we do want to do --- if our measure developer is here, we do want to see if you can tell us one more measure.

MS. DAVIES: This is Sheryl Davies. We're here with AHRQ. Can you verify for us that you will not be doing the ad hoc measures today?

CO-CHAIR SEPTIMUS: That's correct.

MS. DAVIES: Thank you.

CO-CHAIR SEPTIMUS: Appreciate your patience.

Now, Josh is going to present this so I don't want you to influence -- Jason. Jason is going to do this so I don't want you to influence your results.

MS. ARDIZZONE: I just want to let you know I'm going to leave at 3:00 so I don't know what quorum, what it needs to look like or --

CO-CHAIR SEPTIMUS: There's 20 minutes.

MS. ARDIZZONE: Okay. And I do want

Jason to get his full, not hurried.

DR. BURSTIN: If nothing else, it might be useful to at least allow Jason to present so people can hear it even if we don't run through the evaluation. I think it's easier if you go with the asking questions in person. It's always more awkward on the phone. So, even if you're not finished with the evaluation you can do that, I think. And then make sure to save a few moments for public comment for sure.

DR. ADELMAN: Okay, sure. I'm going to start.

All right, thank you everybody. On the phone with me are two folks from the Montefiore team, Dr. Will Southern who is the Chief of the Division of Hospital Medicine for Einstein and Montefiore who collaborated with me on the development of the measures as well as the research we've done with it. And Dr. Brandon Young who is a statistician and the Senior Epidemiologist from Montefiore.

So, I have some slides. So, if you

look forward, I'm going to explain the measures very clearly and quickly.

So, first, some of these slides have some animation, so I'll just tell you next and if you could advance. I'll tell you when. Okay. All right.

The measure very simply, it's looking for an automated method for identifying wrong patient errors. It looks for when a doctor places an order as you can see on the slide in front of you and then cancels the order. Next.

Within 10 minutes and then immediately after, same doctor orders the exact same -- or not the exact, places the same order on a different patient. Next.

So, the programmer who helped develop it used to call it the oops career, like oops I'm on the wrong patient, let me catch it and fix it. We call it the retract and reorder measure. Next slide.

So, prior to our work, the most wrong patient errors anybody ever identified was around

nine or ten by voluntary reporting. You could see when we developed this, we found 6,885 in one year. And even though I showed you we set a limit for 10 minutes for retraction, on average it was retracted within one minute and eighteen seconds. So, if you track yourself it's usually pretty quick. Next.

To validate this, we got IRB approval, got a precision analysis and we called 243 people in near real time, shortly after they made the error, and we asked them was it in fact a wrong patient error. And 170 or 76.2 percent confirmed it was. Next.

So, if you correct for that it was still over 5,000 in one year. Fourteen a day. One out of every six providers. Next.

So, validity. First of all, I just want to point out that this is a unique measure. We talked earlier about creative new measures. It's very different than almost everything I think we've seen in the last two days and that is measuring actual actions of providers. It's not

what somebody charted and then a chart reviewer extracted or abstracted. It's not the doctor's documentation. It's, in fact, the doctor's action. Like you must place an order, and then it has to see that that order was actually canceled. And that allows for some special things.

So, first of all, we hardly heard today before the ability to call people an hour after they made an error and said, did you just make an error? So, when we validated we really validated. Next.

And it uses very basic data. I mean, there's nothing more fundamental than who placed an order, what time that order was placed, and what was the order and then was it DC'd. That's all the data that it uses. Next.

And I guess I made this point. There's no human interpretation. It's just actually what happened. There's no ICD-9 codes. No chart retraction. No voluntary reporting. Next.

So, I cut and pasted from, you know, I guess I have the manual that we get, too, so I cut and pasted some of the rules for validation. It says another authoritative source for the same information. So, as I said, we got to call people in real time and positive prediction of 76.2 percent with a very narrow competence interval plus or minus five percent.

I didn't put this in the packet, but I saw some of the comments about reliability and validation and the VA in New York, I gave them the exact instructions of the measure. I gave them my IRB protocol. I gave them my script for making phone calls. And they made not as many as us. They made 35 and 26 out of 35 so their positive predictive value was very similar to ours. It was 74.3 percent.

And then another hospital used chart review. Now chart review is different because in that case, they look for -- if somebody orders insulin on a patient and cancels it and then orders it on another patient, if the first

patient has no diabetes, no insulin and then they cancel it and then the second has diabetes and insulin, they decide it's a good chance it was a wrong patient error. But if by chance both had diabetes, then it was hard to determine. So, you'll see by chart review, they looked at 200 charts, 61 percent were valid, but another 38 percent were indeterminate. So, it's probably more than 61 percent and in the range of what saw in the other slides. And each one of these was done at a different hospital with a different EMR. Next.

So, face validity, a lot of experts have endorsed this. So, yesterday you heard the chair of the NQF Health IT Committee. He and David Bates also wrote a letter that's in the folder there. Actually there's a JAMIA article that suggests that we submit these measures to ONC on a quarterly basis. Actually, ONC has the SAFER Guides which you could all see on line, and they recommend using the measure now. And then, of course, the Health IT Committee as I

mentioned. That's why I decided to submit it. I'm also on that, and NQF Health IT Safety Division, and they kept using this as an example of a good health IT safety measure, so we decided to submit. Next.

Also it says, performance scores resulting from the measure as specified can be used to distinguish good from poor quality. Well, we've been using this measure a lot and publishing a lot to test intervention. It was not possible to do any of the work that we would do beforehand because hospitals would have like seven errors. You couldn't test because there wasn't enough, and they were voluntary reported.

When you have 5,000, you can do randomized controlled trials, which we did, and we showed that alerts decreased errors. We showed that if you type in initials, age and vendor that decreased errors. We just got a paper accepted in pediatrics that shows that when you name children, baby boy, baby girl temporarily, that that increases errors, and if

you use the mother's first name like Wendy that decreased errors. And we're funded by AHRQ to study how many records are in our state. So, we've been using it successfully a lot to show that we can decrease these errors. Next.

So, I said this already. As far as reliability, you know, some of the rules are, they have to be well-defined and precisely specified, and it uses very basic data, and I showed you with the graphic at the beginning. It just looks at when orders are placed, canceled. Next.

And it asks if it can be consistently implemented, and we think it can be for the reasons we said, and I'll show you, we've been sharing it. But also our system that we tested with GE. We're going to Epic, and we built it in Epic, but we haven't yet, but when we're ready we're just going to share it with the entire Epic community which, you know, Epic says, well 50 percent of the patients in the United States are going to have it, and we can just share it.

Next.

So, this is the exact quote, if you look under 2a. Reliable testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population. So, essentially, if you re-run it, will you get the same results? So, if it's chart extraction of doctor's notes, and doctors document differently and the chart inspectors read it differently. our infection control preventionists are trying to figure out if there's a CAUTI. You know, the inter-rater reliability may vary. Here, it's computers pulling direct actions of doctors. So, it's just very -- and it was hard to describe how to do the reliability, so I consulted with Andrew and others at NQF, and initially we just said, it's reliable for these reasons. Next.

And these are some of the options that NQF offers. You can to inter-rater reliability, but there's not people extracting, so how do you do it? You could do test, re-test. Next.

What we did in the end is we just ran it. We did test, re-test. We ran it a couple of times to show that because it's using this data, it always, you know, repeats itself almost identically, and if you compare, the comparisons are identical. Next slide.

We shared it with many hospitals, all on different EMRs and, you know, we give them the requirements. It's anywhere from a couple of hours to a couple of days to write the brief because it's relatively simple, and the results are remarkably similar. They're not exactly the same because the EMRs are different. Some allow for a couple of records open. Some, you know, have different fonts, patient photos. So, that's the whole point of all this is that we need to see the safety of EMRs and then push them to put in photos and limits of one and different things to have protection. Next slide.

So, this was, and I mentioned this before, but before our research, the most errors that I could find that were quantified was in the

MedMARx database, 120 facilities. Checked off that CPOE was the cause, and they found a mean of 9 wrong patient errors, and then next slide.

I found one study that did a large chart review, and they found many errors, but only two were wrong patient errors. So, with voluntary reporting and chart review, you just can't ascertain what's going on. Everybody knew it was a problem, but it wasn't until we came up with the automated measures that we could do all the research that we did that led to all the national recommendations. Next slide.

So, I'm almost done. So, all of this started in 2011, the Institute of Medicine came out with a report about Health IT and Patient Safety. That led to AHRQ funding more grants, which I got funded by NQF to make other HIT Safety Committee. Next.

So, ONC made a HIT Patient Safety Action & Surveillance Plan. As I said, AHRQ, ONC, and they're all recommending we need Health IT Safety Measures, and then they're, you know,

highlighting this one as the first, really, as an example.

And as I mentioned the SAFER guides are already recommending on line that people do this. Next.

And then there's all these letters of support. Next.

So, I kind of went over scientific acceptability and importance, but I think in that that I discussed feasibility like it's basic data. I talked about usability. I think it will be used for more research and for surveillance like David Hunt from ONC was going to call in yesterday but because of the time change -- but in this JAMIA article that several of our Health IT national experts, they said we should send this data quarterly so that people will be compelled to put in photos and have them type in initials and do all these things to keep driving down the rate. Next.

So, I'm pretty much over. It's unique for the reasons I told. We've already done

current use research, and research was not possible for it. And it can be used for surveillance. I think it's valid, reliable. I think it's important, and I think it's feasible. I think it's usable, and there's no competing measures or anything like it.

Thank you.

CO-CHAIR SEPTIMUS: Well, that was an actual incredible job. Outstanding job.

DR. ADELMAN: Thank you.

CO-CHAIR SEPTIMUS: That doesn't mean when we have our phone call, we're going to endorse it, but it was a great presentation, Jason. It really was. I don't know whether or not you'd share that presentation with the committee.

DR. ADELMAN: Oh, yes, sure, we can share it.

CO-CHAIR SEPTIMUS: Yes, that would be great. Put it on SharePoint, yes.

So, we have about five minutes for people to ask Jason questions. We'll go through

the data elements when we have our follow-up phone call, and then we'll take about five plus minutes, and then we have to have public comments.

So, Laura.

MS. ARDIZZONE: Just two quick questions.

So, you presented some data that most people correct their near misses within a minute and 18 seconds. How many people are -- like what is the harm? How many patients are we giving wrong dosage and wrong medications? Because inherent in this is a check and balance system. So, there's a planned intervention to pick up these near misses. There's a nurse check for right patient, right medication, that sort of thing. I think where we're really vulnerable is where there is no check for that near miss.

DR. ADELMAN: Yes, so not as much is known about the actual errors that reach the patient and cause harm. It's, you know, I gave you my very, very quick presentation. I have my

much longer one where I share some data on that question like, for example, ECRI did a study and showed that 15 percent of the adverse events that were reported to their PSO were related to wrong patient errors. And we started this whole thing because we had many real wrong patient errors, and most of the people that have used my measure and shared it have had the same experiences. But they haven't done as good of a job quantifying the errors because it relies on voluntary reporting, and people just don't like to report. That's what makes this measure so powerful. So, it is after near misses which allows for the research to study it, but we still have to use like voluntary reporting to find out the really bad things, and so we have enough to know that they're out there and when they happen they're bad. But we don't have enough information to quantify it.

CO-CHAIR SEPTIMUS: Okay, Josh, then Louie, and then Charlotte and then we'll go to public comment.

DR. APPLEGATE: And then Kimberly, too. I'm on the line. I have a --

CO-CHAIR SEPTIMUS: Well, I'll tell you what. Kimberly, since you've been so patient on the line, Josh, if you don't mind. Kimberly, I didn't know. I apologize. Kimberly, go ahead.

DR. APPLEGATE: Thank you so much. And, Jason, I would love to talk to you offline further about this because I really support the measure without having all the details of it, but I will say that I want to support this by saying that I'm doing my own research from an imaging perspective trying to get photo IDs into the CAHPS record. I know about the Denver Children's Hospital work, and I also know about Hardeep Singh's work at the VA in Texas. And I support what you're trying to do. I think it is somewhat of a connectivity issue and interoperability issue. In that I've done some research at Emory trying to get photos of patients into the record with some publications in the imaging side and at SIIM, the Society for Imaging Informatics in

Medicine, which we presented this year and last year. And I will say that you're right. There's very little wrong patient error research, and we have done some on the imaging side to show that even though there are low rates of it, the consequences can be quite negative for patient safety, and we continue to look at it in our institution. And when the technologists have the drop-down list where they can send patient imaging to the wrong folder, nobody can find it. And so we have no idea how often it occurs, and nobody wants to talk about it. And I would just strongly support this work, and I'm trying to get support for more research on the imaging side. And as we do millions and millions and millions of imaging tests, and we continue to ask our workers to do more and more, you can imagine that even one of these cases can lead to very bad consequences. And we are finding that we may have a joint commission, two-patient identifier requirement. But when we have many demands on our technologists to hurry up and do the test,

they do not follow that requirement, particularly on our portable studies, and that's our most common event that we have where they're not following it. And you can imagine that they are low on the totem pole, and when people are demanding they take the image without double-checking, this happens more than you think.

So, I commend Jason for doing this research.

Thank you.

CO-CHAIR SEPTIMUS: Thank you very much. So, we will be quick so we can everybody's answers in before we have to go to public comment. So, Josh is next.

DR. RISING: Great. Very quick question.

So, when you called the quarter of the people, right, you know, you determined it was not kind of an error. So, what so those 25 percent of people say that, you know, convinced you that, yes, this was not a kind of erroneous order in the first place?

DR. ADELMAN: Sometimes people are placing a couple of orders in bulk like we just talked about Coumadin. At the end of the day before they go home they check the INR, and they order Coumadin. So, they order it for the first patient. They order it for the second, and then they realize, oh, the attending said, don't order Coumadin because the patient is going to the operating room. So, they order it, and they cancel it, and then they order on the third. So, it was a type of error. It just wasn't a wrong patient error. But it looks like a retract and reorder event. A false positive.

CO-CHAIR SEPTIMUS: Lillee, do you still want to make a comment? Yes.

MS. GELINAS: First of all, congratulations, really. I look forward to when we have the phone call to vote.

But my question has to do when you talk about spread and interoperability. We're a large system including internationals. So, we have to report measures, not only domestically

but internationally. Our two largest systems are Cerner and MEDITECH. I find when we try to introduce new measures, our new reporting mechanisms, Cerner is much more amenable than MEDITECH.

Have you quantified the financial burden to systems who have to pay for the -- I would say the updates. I forgot how MEDITECH puts it, upgrades, but I just find that that is a piece when we move to really important components like this, we just run into some buzz saws with the vendors that haven't been quantified so it's really hard to sell to the CMIO and CIO.

That said, we do have Midas. And so we have a good event reporting system, but I know we don't have anything like this. Have you also thought about anyway we can be collecting this through some of our normal mechanisms?

CO-CHAIR SEPTIMUS: Josh hold that. We have to get public comment before 3:00, and then we'll come back. I hate to do this but, no.

So, public comment. Operator?

OPERATOR: If you would like to make a comment please press star, then the number 1.

There are no further comments from the phone line.

CO-CHAIR SEPTIMUS: Thank you, Operator.

I'm sorry about that.

DR. ADELMAN: That's okay.

One of the questions was, you said there were two. One of them was about the cost. So, one of the slides I had there showed multiple systems that we implemented. We started it in GE, moved it to Epic. There's a Cerner. Allscripts is there. The VA system is there. The Brigham's home-grown system is there. And what I can tell you is, it uses -- in none of those systems did we actually work with the vendor meaning like all of them allow you to get the very simple reports of like medications and orders. And so the data comes out, and then a programmer writes a query just to see was an order placed, retracted, and so it's really a

report against the data as opposed to like some code implemented within the vendor. And so I can tell you, you know, we shared it with Dr. Schreiber before coming here because she was reviewing it, and she said, hey, how hard is this? And we gave her the requirements and I think it was like two hours of writing it, and so it was not difficult at all.

CO-CHAIR SEPTIMUS: I want you to know that we feel your pain since we're MEDITECH shop also.

Yanling.

DR. YU: I just have quick question. I really applaud your efforts to have this measure. It's really cool.

DR. ADELMAN: Thanks.

DR. YU: Very creative. You have a diagram that basically says yes or no, if adequate human recover, it will go to a near miss. If not, it goes to an adverse event.

DR. ADELMAN: Yes.

DR. YU: So, we all know that adverse

events sometime are under-reported or hard to define, so they're not reported as an adverse event.

Do you have any estimate how that would have inflated your estimate or near miss?

DR. ADELMAN: I mean --

DR. YU: Is the error bar, you know -- or affect your error bar for near miss?

DR. ADELMAN: I think I understand your question.

That diagram I took from the Institute of Medicine report. I think it was just called "Patient Safety." And, you know, that's a seven-year old diagram about near misses. And everybody believes in near misses. NQF has another committee where they do common formats, and we report adverse events and near misses.

Even though the evidence connecting adverse events and near misses is not as strong as we'd like. A lot of it comes from the transportation industry. There's some in health care. And there are these rules of thumb like

for every actual adverse event there's 100 near misses, so you can make some estimates. I don't know how well they hold up. So like everyone that deals with patient safety believes in the value of near misses because there certainly are many more of them, so you can study things that will then ultimately prevent the actual event. I'm not exactly sure of the proportion of these self-caught near misses and the actual errors. I just know, you know, we started the whole thing because we were having too many people getting hurt from wrong patient errors, you know. It's sort of like, you place the order on the wrong patient, and then by the grace of God you realize, oh, crap, I'm on the wrong patient, and you cancel it. And if that happens which obviously happens a lot, then it doesn't, you know, reach the patient. And if you don't have that thought it's not like a pharmacist checking it or a nurse checking it. It's you in the back of your mind realizing, oh, crap, I'm on the wrong patient. It's not a very reliable safety

mechanism. And we saw that too often they were reaching the patient.

CO-CHAIR SEPTIMUS: Charlotte.

DR. ALEXANDER: Well, first, I want to commend you. This is really needed. It's outside of the box. It's really great.

My angst is that when, from a patient safety point of view, we look at near misses as really great opportunities to learn. And it's super, super, super important that there's no punitive attachment to it. My concern is if this rolls out as a measure, and hospitals are then being measured by that, and then there's a pay-for-performance problem with it, we're just doing all the wrong stuff with a near miss.

DR. ADELMAN: Yes.

DR. ALEXANDER: And so this is one that is perfect to use, to learn, to develop new things, to --- if we were to counter and catch, but I really hate to see near misses go down that track of being punitive.

DR. ADELMAN: So, I taught Just Culture

at the National Patient Safety Foundation Annual Congress the last two years in a row, so I'm a big, you now, proponent of Just Culture and differentiating human failure from system failures. And this is all about system failures. All the studies that have come from it taught us how to make the systems safer. And to your point, you know, one of the worst errors that we had where, I think we're in a public comment, right, so I won't say what happened at Montefiore, but it was bad and --

CO-CHAIR SEPTIMUS: No, we're past public comment.

DR. ADELMAN: We're past, but I just don't know if they're --

CO-CHAIR SEPTIMUS: Right.

DR. ADELMAN: We had a bad wrong patient that would really scare you, and at the time we had a root cause analysis, and the theme was, the intern must know what patient you're placing orders on. And then we did a study, and we showed everybody places orders on the wrong

patient and it's not the intern. It's the system. And then we did all this research to show if you have a pop-up that will decrease errors. If you have them re-verify that they're on the right patient with the initials, that will decrease errors, and now even if you change the name of children, so you don't call every kid in the NICU, baby boy, baby girl, that decreases errors. So, I think, you know, the point is about making the system safer. And I guess what we're trying to do is hold like vendors, not individuals but vendors accountable and health systems. It's time to invest in patient photographs. You know, it's time to -- and I think that will be the level of --

CO-CHAIR SEPTIMUS: And, Michelle.

DR. SCHREIBER: Thanks. As you know, you and I have worked together in reviewing this, and we actually did try it at the Henry Ford Health System, so just a couple of comments.

I know Lillee is gone, but this is entirely scalable. It took us only a couple of

hours to do it, and there's actually virtually no cost to organizations in doing it. And there are great lessons to be learned.

Wrong medication errors are almost becoming endemic now with CPOE, and so I think this is a really creative measure, and not only that: you thought of how to do it from the very beginning, so I just want to congratulate you.

Just one comment and a little bit of a question. In some of how you looked at this, it was all orders, and some of how you looked at this and verified it was medications only. So, if you can just clarify, is this always going to be all orders or medication orders only?

DR. ADELMAN: Right. We did all orders and then parsed it by all types of orders in our JAMIA paper, which I think is attached to the application. We showed how wrong patient errors by medication orders, radiology orders, nursing orders, and then there was a table where I showed many different hospitals that implement it. Some like the Brigham, for example, was funded by the

FDA to do a medication safety project so they ran it with a particular interest on medications. And it wasn't that they couldn't do it on all, and I actually didn't go back and ask them. I suspect that they did it for all and then only reported the medications as part of that project. And there was another hospital that similar story.

I think it's intended and meant and should be used for all orders, but you can parse it any way you want.

CO-CHAIR SEPTIMUS: Yes, Jason and I talked last night as we walked back in the rain.

Another thing to look at. I mean I have been doing CPOE for a long time, and either I have not recognized it but I've not had that opportunity to put it on the wrong patient. I'm sure that it can happen, that's for sure. You proved that to me. But what I see more often is not that the order is placed on the wrong patient, but that the order is wrong. So, you asked for a specimen from one site, but you

really meant another site. And I wonder whether or not you might want to look at wrong order on the right patient.

DR. ADELMAN: Yes, so interesting that you say that because, you know, all this work, I mean, for us it stems from actual wrong patient errors. But the fact that NQF has a Health IT Safety Committee and AHRQ, Congress gave AHRQ an extra $4 million to study Health IT Safety. And I just submitted two weeks ago a new grant to AHRQ saying that we want to expand this measure. The concept of retract and reorder to find other types of errors.

So, just very quickly, an example of a wrong order would be a wrong medication would be instead of looking for orders that are placed on a patient, retracted and placed on another patient, we took the IMCP list of similar sounding medications and looked for when somebody orders Clonidine on a patient, cancels it very quickly and then on the same patient order Klonopin, which often gets confused. So, it's

wrong drug and, you know, because I sit on this committee, and I sit on the Health IT Safety Committee, I sort of made the case in the funding opportunity in the proposal that we want to develop the measures, test them, and I'm intentionally testing in such a way that it will be -- we can demonstrate validity and reliability to the NQF so it will be ready for endorsement.

So, if I get funded I will be back in a few years with, you know, more Health IT Safety Measures to address just what you asked.

DR. BURSTIN: To add in, you know, as we are finishing up this work on Health IT and Patient Safety, we'll also make sure we send that draft report to all of you because part of what you -- you remember Andrew from the last project. Andrew has been leading this work and so we've got -- they came up with a list of about 100 concepts that potentially could be measures that reflect Health IT and Safety. So, we'd love to get your input on that. There's a long list of them, Michelle, beyond this one. This was the

first one that at least we knew there was something in development and at least get a chance to look at it.

DR. ADELMAN: Oh, can I please thank Andrew in the back for helping me and guiding me and the NQF staff, they were super helpful.

CO-CHAIR SEPTIMUS: And so, yes, we should end the meeting by, first of all, thanking the NQF staff for the support they provide this committee. I mean, we know how much work it takes to even get to this point. I know we're not finished with their work, but they deserve a really big round of applause.

There will be a follow-up phone call as you know. I think there's going to be lots of work done between now and whenever that phone call is going to take place. And I can just say I'm just amazed at the talent that's in this room. You all make me feel awfully humble and awfully inadequate because of all the talent that's in this room. And I think that Iona may have had to leave early, but I want to thank Iona

for -- we sort of support each other and made sure we don't make too many mistakes. So, we thank you from our perspective for being such a great committee. And I know our work is not done so safe travels to everybody.

DR. BURSTIN: Absolutely. Thanks from us, too.

(Whereupon, the above-entitled matter went off the record at 3:12 p.m.)

H

M

T

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Patient Safety Standing Committee

Before: NQF

Date: 06-18-15

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