NATIONAL QUALITY FORUM

+ + + + +

BEHAVIORAL HEALTH STEERING COMMITTEE

+ + + + +

TUESDAY
APRIL 17, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:30 a.m., Peter Briss and Harold Pincus, Co-Chairs, presiding.

PRESENT:

PETER BRISS, MD, MPH, Co-Chair
HAROLD PINCUS, MD, Co-Chair
CAROLINE CARNEY-DOEBBELING, MD, MSc, Medical Officer, MDwise, Inc.
MADY CHALK, PhD, Director, Treatment Research Institute
DAVID EINZIG, MD, Children=s Hospitals and Clinics of Minnesota
NANCY HANRAHAN, RN, PhD, University of Pennsylvania
EMMA HOO, Director, Pacific Business Group on Healthcare
DOLORES KELLEHER, MS, DMH, Principal, D. Kelleher Consulting
PARINDA KHATRI, PhD, Director, Cherokee Health Systems
TAMI MARK, MBA, PhD, Senior Director, Thomson Reuters Healthcare, Inc.
BERNADETTE MELYNK, RN, CPNP, PhD, Dean, The Ohio State University College of Nursing

MADELINE NAEGLE, APRN-BC, PhD, FAAN,
Professor, College of Nursing, New York University*
DAVID PATING, MD, Chief, Kaiser Permanente Medical Center
KARLENE PHILLIPS, BSN, RN, Director, Mayo Clinic Health System
VANITA PINDOLIA, PharmD, HFHS/HAP
Vice-President Ambulatory Clinical Pharmacy Programs, Henry Ford Health System
JEFFREY SAMET, MA, MPH, MD, Chief, Department of Medicine, Boston University
LISA SHEA, MD, Associate Medical Director, Butler Hospital, Providence, RI
JEFFREY SUSMAN, MD, Dean, Northeast Ohio Medical University
LYNN WEGNER, MD, Clinical Associate Professor, UNC Department of Pediatrics
BONNIE ZIMA, MD, MPH, Professor-in-Residence, UCLA Department pf Psychiatry and Bio Behavioral Sciences

NQF STAFF:

HELEN BURSTIN, MD, MPH
SARAH FANTA
ANGELA FRANKLIN, JD
ANN HAMMERSMITH, JD
SARAH LASH
EVAN WILLIAMSON, MPH, MS

ALSO PRESENT:

DAWN ALAYON, National Committee for Quality Assurance
MARY BARTON, National Committee for Quality Assurance
MICHAEL FIORI, The Joint Commission
ERIC GOPLERUD, The Joint Commission

JEREMY GOTTLICH, National Committee for Quality Assurance
NANCY LAWLER, The Joint Commission
KATHLEEN McCANN, National Association of Psychiatric Health Systems
STEPHANIE MIKA, U.S. Department of Health and Human Services
STEVEN SCHMALTZ, The Joint Commission*
SAMANTHA TIERNEY, Physician Consortium for Performance Improvement
ANN WATT, The Joint Commission

*Participating by teleconference

A G E N D A

Welcome and Introductions..... 6

Disclosures of Interest..... 11

Government Efforts Around Behavioral Health..... 26

Project Introduction and Overview of Evaluation Process..... 34

Measure 1651: TOB-1 Tobacco Use Screening..... 74

Measure 1654: TOB-2 Tobacco Use Treatment Provided..... 115

Measure 1656: TOB-3 Tobacco Use Treatment Provided or Offered at Discharge..... 174

Measure 1657: TOB-4 Tobacco Use: Assessing Status After Discharge..... 213

Measure 0028: Measure Pair: Tobacco Use Assessment and Cessation Intervention..... 236

Measure 0027: Smoking Cessation, Medical Assistance, Advising Smokers to Quit, Discussing Smoking Cessation Medications, Discussing Smoking Cessation Strategies..... 262

NQF Member/Public Comment..... 284

Related and Competing Measures Discussion..... 285

Measure 0004: Initiation and Engagement of Alcohol and Other Drug Dependence Treatment..... 312

Measure 1661: SUB-1 Alcohol Use Screening..... 286

Measure 1663: SUB-2 Alcohol Use Brief Intervention Provided or Offered; Alcohol Use Brief Intervention..... 381

Measure 1664: SUB-3 Alcohol and Other Drug use Disorder Treatment Provided or Offered at Discharge, Alcohol and Other Drug use Order Treatment at Discharge..... 397

NQF Member/Public Comment..... 422

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

8:32 a.m.

Welcome and Introductions

CO-CHAIR BRISS: So good morning. Good morning, welcome. So in the room, a lot of people are still fiddling with computers, but while we do that, I thought I'd welcome everybody to the meeting, and so I'm Peter Briss.

I'm the medical director in the Chronic Disease Center at CDC, the Centers for Disease Control and Prevention in Atlanta, and it's my honor to get to co-chair this merry band this morning with Dr. Harold Pincus.

CO-CHAIR PINCUS: So welcome, everybody. We're going to get started very soon, but we want to have a time to introduce everybody, and also a time to give you a sense of the way the day is going to work out.

This is a complicated process. It's deceptively complex, and at times we have to go through all the criteria, all the time for each of the different measures.

So we can sometimes seem redundant, but it's important for the staff and for the evidence based behind what we're doing to actually go through that and have some serious discussions about each of the issues.

We're going to try to go through it as efficiently as possible, so that ideally we will not waste a lot of time. On the other hand, we do want to hear everybody's views, and so we want to make sure that everybody has a chance to give their views on each of the measures and, for that matter, on each of the criteria for each of the measures.

So we're going to try to be as fair as possible and try to identify people that want to speak, in some systematic way. I don't know if you know, but what has been used before is for people who want to speak to put their card up like this, and ideally if they could do it in a way where Peter and I could see it, because at least I can't remember everybody's name. I have a hard time sometimes remember my kids' name.

CO-CHAIR BRISS: So I think we want to do introductions and conflict statements next, and I think that Ann will walk us through the next part of the agenda.

MS. HAMMERSMITH: Good morning everyone. I'm Ann Hammersmith. I'm NQF's general counsel. This is the part of every committee meeting where we combine introductions with the conflict of interest disclosures.

If you recall, probably several months ago you should have received a form from us, where we asked you some specifics about your work, about outside activities and so on.

What we want to do today, in the spirit of transparency and openness, is to just go around the table and have you disclose anything that you think your fellow Committee members should know about your activities.

I want to remind you that the fact that you disclosed something does not mean that you have a conflict of interest. It's simply a

disclosure.

We certainly don't expect you, nor do we want you to frankly recount your entire CV, because that would take up the entire meeting. You also don't need to cite every single thing you put down on your form, unless you think it's relevant.

I want to remind you about a few other things before you start the disclosures. We're particularly interested in your disclosing any consulting research or grants that you have, that you believe are relevant to what's before this Committee today.

I also want to remind you that you sit as an individual. We often have members who, when they make their disclosure, will say I am Joe Smith and I am here representing the American Association of, fill in the blank.

You may be employed by the American Association of fill in the blank. The American Association of fill in the blank may have nominated you, but you are not here representing

their views. You are here because you are an expert, and you sit as an individual.

Finally, I want to remind you of one other thing. We often have Committee members go around the table and say "I have no financial conflict," or "I have no financial disclosure."

A financial conflict of interest is of course important, but in this world we also look at what you've been involved in and money may not have changed hands.

You may have served on some guidelines committee or something like that that may be relevant, something that you want to disclose. No money may have changed hands, but it's still something that it might be appropriate for you to reveal.

So what that, I'm going to have you go around the table, tell us who you are, where you work, if you have any disclosures, and I'll start with the chairs.

Disclosures of Interest

CO-CHAIR BRISS: And we got very

specific instructions from staff to model brevity. So again, I'm Peter Briss. I'm the medical director in the Chronic Disease Center at CDC, and I've done a lot of extensive work with every conceivable committee and subcommittee at NQF lately, and I have no conflicts.

CO-CHAIR PINCUS: Yeah. Part of the problem is that I was at a different meeting where the speakers worked differently, but so I'm Harold Pincus. I'm a professor at Columbia University, and I'm vice chair of the Department of Psychiatry there.

I also have a role as co-PI of the Irving Institute for Clinical and Translational Research, and as the director of Quality and Outcomes Research for New York Presbyterian Hospital. I'm also adjunct staff at the RAND Corporation.

My research that I do is mostly health services research, mental health services and policy research, that's all been funded by not-for-profit organizations,

including the government.

But I have some roles that I have as a consultant, and I've been a consultant with Mathematica and with the Altarum Institute and with Manila Consulting, as well as I am on an advisory board for Value Options, but receive no compensation for that.

And also, I'm on an advisory board for the National Committee on Quality Assurance, and I also am on the board of the American Society for Clinical Psychopharmacology.

(Off record comments.)

CO-CHAIR PINCUS: Well, I've been involved in the development of some of the measures that the RAND Corporation developed, as part of an effort, a project to evaluate the quality fo mental health care at the Veterans Administration.

DR. BURSTIN: But they're not submitted.

CO-CHAIR PINCUS: But they're not submitted. They're similar to but they weren't

submitted. They were actually submitted as part of a different process, that then apparently NCQA took them up and submitted them.

DR. EINZIG: I'm David Einzig from Minnesota, St. Paul, Children's Hospitals and Clinics of Minnesota, based in St. Paul. I'm a child psychiatrist and a pediatrician. I did the training in the Triple Board program. I'm president-elect of the Minnesota Society for Child and Adolescent Psychiatry. No disclosures and nothing else.

DR. CARNEY-DOEBBELING: I'm Caroline Carney-Doebbeling. I'm the chief medical officer of MDwise, Incorporated, and previously served as the medical director for the Indiana Medicaid Program. I have no disclosures.

DR. SAMET: Good morning. Jeffrey Samet from Boston, Boston University, professor of Medicine there, chief of general internal medicine. So I'm a general internist by training. I don't think I have a lot of

disclosures.

I am president of the American Board of Addiction Medicine currently, that trains, is our same level training for addiction physicians, and being funded by NIH a number of different studies, one that we actually looked at brief intervention for alcohol, and currently for drug in medical settings.

DR. CHALK: I'm Mady Chalk. I'm Director of Policy Research and Analysis at the Treatment Research Institute. In fact, I was working addictions treatment and performance measurement and policy.

I'm also on the board of the Washington Circle, which looks at and does some consensus processes with regard to measurement. I have no conflicts.

DR. SUSMAN: I'm Jeff Susman. I'm the dean of the Northeast Ohio Medical University, and since going over to the dark side, I have no time for research or anything substantive but passing papers around, and I

have no conflicts.

(Off record comments.)

DR. SUSMAN: Medicare and what CMS says.

(Off record comments.)

DR. SHEA: I'm Lisa Shea. I'm the associate medical director of Quality and Regulation at Butler Hospital in Providence, Rhode Island. I also am a trustee for the National Association of Psychiatric Health Care Systems, and I don't have any conflicts.

DR. MARK: Hi. I'm Tami Mark. I'm a senior director at Thomson Reuters. My training is as a health economist, behavioral health services researcher.

My standard disclaimer is that Thomson Reuters provides information assets and consulting services to all aspects of the health care system, employers, health plans, pharmaceutical companies, the federal government, providers, hospitals, etcetera.

DR. WEGNER: Good morning. I'm

Lynn Wegner. I'm division chief of the Developmental Behavioral Pediatrics Program at UNC in Chapel Hill. I'm representing the American Academy of Pediatrics for this project.

I have a variety of liaison appointments to the American Psychiatric Association and the American Academy of Child and Adolescent Psychiatry.

Probably my big claim to fame is that I'm on the Committee on Coding and Nomenclature for the AAAP, and I am extremely passionate about financing and the lack thereof within the system, and if anybody wants to talk to me, I would be glad to.

MS. PHILLIPS: Hi. I'm Karlene Phillips. I'm the Director of Behavioral Health Inpatient Services --

(Off record comments.)

MS. PHILLIPS: --at Mayo Clinic Health System in Eau Claire, Wisconsin.

DR. HANRAHAN: My name is Nancy Hanrahan, and I'm not connected to the Internet,

if someone would like to help me.

(Laughter.)

DR. HANRAHAN: I'm an associate professor at the University of Pennsylvania. I'm a psychiatric nurse, and I've been doing in this business a long time, both as a clinician, researcher, administrator, whatever, but no conflicts of interest.

DR. KELLEHER: I'm Dolores Kelleher, Dodi to people who know me, and I am an independent consultant in the areas of health, wellness and value-based design, primarily to employers or employer groups, and previously I spent many years with United Behavioral Health, doing behavioral health program design and implementation.

Worked for the last few years for Safeway, building out their health and wellness strategy, and as far as I know, I have no conflicts.

MS. HOO: I'm Emma Hoo with the Pacific Business Group on Health, working on

care system redesign. I have no conflicts.

DR. PINDOLIA: Hi. I'm Vanita Pindolia with Henry Ford Health System. I've been on the -- I'm the Vice President of Ambulatory Clinical Pharmacy Programs, so I develop also the transition of care programs from the hospitals, and work with the case managers for those programs for medication issues.

The only potential conflict, but it's really not, it's just that I'm also on another committee, and it's with URAC, and it's the Measurement Advisory Committee. They aren't doing any behavioral med. I'm not on the group that's with behavioral medicine. I oversee the diabetes measurements, but that's about it.

DR. KHATRI: I'm Parinda Khatri. I'm Director of Integrated Care at Cherokee Health Systems. We're a comprehensive community health care organization providing integrated medical primary care, behavioral

health and substance abuse services in Tennessee.

DR. MELNYK: Good morning. I'm Bern Melnyk, and I'm the University's Chief Wellness Officer and Dean of the College of the Nursing at the Ohio State University. I am currently conducting NIH-funded research, cognitive behavioral, healthy lifestyle interventions with high school adolescents.

I just came off of a four year term on the United States Preventive Services Task Force, and I'm happy to be here. No conflicts.

DR. ZIMA: I'm Bonnie Zima. I'm a child psychiatrist and health services researcher at UCLA. I do receive research money from the National Institute of Mental Health, and I'm also an investigator on the AHRQ CMS Center of Excellence based at the University of Washington, through a subcontract at RAND.

DR. PATING: I'm David Pating. I'm region chair of Addiction Medicine for Kaiser Permanente Northern California, where I'm also

a co-PI on a NIAAA screening brief intervention study. I'm a member of the board of the American Society of Addiction Medicine and a mental health commissioner for the state of California.

MS. MIKA: Hi. My name is Stephanie Mika. I am here representing the Office of the Assistant Secretary for Planning and Evaluation at HHS, and I am in fact representing the Office of the Assistant Secretary for Planning and Evaluation at HHS.

ASPE is the lead office at HHS that's funding a large body of work with NQF, and this is one of the projects that's there. So I'm here on behalf of the Department, and happy to be here. No conflicts.

MS. HAMMERSMITH: Okay. I understand that there are a few members on the phone.

DR. BURSTIN: Yes.

MS. HAMMERSMITH: Okay. So I'm going to call your names. Is Colleen Barry on the phone? Colleen Barry.

(No response.)

MS. HAMMERSMITH: Michael Lardiere. Michael Lardiere.

(No response.)

MS. HAMMERSMITH: Okay. I'm not doing very well so far. David Mancuso. David Mancuso, are you on the phone?

(No response.)

MS. HAMMERSMITH: Madeline Naegle.

DR. NAEGLE: Madeline Naegle, good morning. Yes, I'm on the phone. I'm a professor at NYU's College of Nursing, a member of the expert panel of the American Association of -- the American Academy of Nursing's expert panel on Substance Abuse and Psychiatric Mental Health Nursing.

Here at NYU, I coordinate our substance-related disorders content, and I'm co-investigator on a NIDA-funded project, Substance Abuse Research, Education and Training, and also a psychotherapist in private practice. No conflicts.

MS. HAMMERSMITH: Okay, thank you. Based on the disclosures this morning, do any of you have any questions of me, or do you have anything that you want to raise with each other, based on the disclosures?

(No response.)

MS. HAMMERSMITH: Okay, thank you. Have a good meeting.

DR. BURSTIN: Just add my welcome. Helen Burstin. I'm the senior vice president for Performance Measures at NQF. It's a pleasure to have you all here. I know many of you. Thank you, for those of you who've been with us before. It's good to have some folks with some experience.

You will notice as you go through the process our criteria continue to get more and more, I think, precise. So even those of you who may have been on the committees in the last year or two will see there's a level of precision.

We hope that helps. We've really been trying to ensure consistency across

steering committees, across projects. So as Angela goes through those, we'll go through that with you, and I have to just add that, you know, if you think I turned out okay, it's all because Jeffrey Samet was my chief resident when I was an intern.

(Laughter.)

DR. NAEGLE: Great.

MS. FRANKLIN: I'm supposed to be modeling using the microphone. I'm Angela Franklin, senior director for the Behavioral Health Project.

MS. FANTA: Hi everyone. I'm Sarah Fanta. I'm the project manager on this Behavioral Health Project, and really looking forward to working with all of you today.

CO-CHAIR BRISS: And could we please introduce the people around.

MR. WILLIAMSON: I'm Evan Williamson. I'm a project analyst. Looking forward to working with you all over the next two days.

CO-CHAIR BRISS: And now let's introduce the people around the sides of the room.

MS. ALAYON: Hello. My name is Dawn Alayon. I'm with the National Committee for Quality Assurance, and I'm a senior health care analyst for NCQA. I will be presenting today on the smoking measure.

PARTICIPANT: My name is Gur Madum (phonetic). I am researcher in the Office Health IT Quality. I'm just attending the meeting. I am not a member.

DR. GOPLERUD: Eric Goplerud, senior vice president, NORC at the University of Chicago, and co-chair of the Joint Commission's Technical Advisory Panel that developed the substance use and tobacco measures.

DR. FIORI: Good morning. Michael Fiori, Professor of Medicine at the University of Wisconsin's School of Medicine and Public Health.

I also co-chaired the Technical

Advisory Panel, and chaired the United States Public Health Service Clinical Practice Guideline panel that produced guidelines for treating tobacco dependence in 1996, 2000 and 2008.

MS. LAWLER: Good morning. I'm Nancy Lawler from the Joint Commission, one of the measure developers.

MS. WATT: Hi. I'm Ann Watt. I also am from the Joint Commission, and --

(Off record comments.)

DR. McCANN: I am Kathleen McCann. I'm the Director of Quality and Regulatory Affairs for the National Association of Psychiatric Health Systems. We're in NQF Provider council members.

MS. LASH: Sarah Lash, NQF staff.

CO-CHAIR BRISS: So we are ahead of schedule, how about that? So Stephanie, you're next on the agenda.

Government Efforts Around Behavioral Health

MS. MIKA: So good morning again,

and thanks especially to Angela and Evan and Sarah for squeezing me into a very packed agenda.

As I said, I'm from the Office of the Assistant Secretary for Planning and Evaluation at HHS, and I wanted to give you a very brief overview of some of the activities going on at HHS, and some of our priorities for this project, so that you have a little bit more of a framework going into the very intense discussion that I know is going to come out of the next few days.

I think some of this will be repetition. I think some of it you've heard from Angela and her team before, and some of it you saw in the call for measures. So forgive me if there's information that you already have here.

As you all know, this two-phased project is intended to endorse individual and composite behavioral health measures that will serve as indicators of quality care access, integration, coordination of care and prevention across all care delivery settings.

This dovetails with a range of activities going on at HHS, including the National Framework for Quality Improvement in Behavioral Health Care, which follows the six priorities that parallel those of the National Quality Strategy, and are based on the IOM's quality reports, and you saw those in the call for measures.

HHS has recognized related opportunities, including supporting the development of a parsimonious set of nationally recognized behavioral health performance measures, that are appropriate at both the national and local levels, expanding cross-agency interests and advancing behavioral health quality measurement, and promoting alignment with the implementation of the National Quality Strategy.

I apologize for my voice. I'm recovering from a cold. So I wanted to touch on a few activities going on across the agency right now, and I hope that Peter will be able to share

some additional information about what's going on at CDC, so I won't touch on CDC today.

So I'll start with HRSA, and Gur (phonetic) can speak more directly to what's happening at HRSA, if you have more questions about that.

But HRSA's priority and programs in the areas related to those activities include FQHCs and primary safety net providers, whose patients cite depression as the third most common reason for a visit; HIV/AIDS, with a behavioral health condition as a possible comorbidity in as many as half of all HIV and AIDS patients; maternal and child health, particularly in the Healthy Start program, which provides case management, depression screening and educational activities, often including formal smoking cessation programs for women in areas with high rates of infant mortality and shortages of health care providers; and workforce programs, such as the more then 3,000 National Health Service core participants, who

provided behavioral health services in 2011.

At CMS, work related to the behavioral health quality measurement activities include the implementation of a new inpatient psychiatric hospital quality reporting program, which I'm sure many of you are familiar with; development of measures on the use of anti-psychotic medications for Medicare patients, including a measure this steering committee is currently considering; and inclusion of a number of behavioral measures for meaningful use eligible professionals and the recent EHR incentive Notice of Proposed Rulemaking, and in the physician quality reporting system.

CMS and AHRQ are also currently working, collaborating on the CHPRA pediatric quality measures program, which over the next several years will include the development of behavioral health measures, with topics including adolescent depression, screening and follow-up; ADHD diagnosis and follow-up; mental

health-focused readmission; medication reconciliation with a focus on mental health medications; and alcohol and substance abuse screening in children and adolescents, I'm sorry, in adolescents.

There's some work going on at ASPE, in collaboration with SAMHSA, to develop measures to assess the quality of care provided to Medicaid enrollees diagnosed with schizophrenia, a number of which are under consideration by this Committee at this very meeting.

And in the next year or two, SAMHSA and ASPE are collaborating to identify, develop and pilot quality of care measures that capture the broad range of needs for adults and children who receive behavioral health services in public systems, and that can help improve the emotional and behavioral well-being of Americans.

Finally, SAMHSA has engaged in work under its strategic initiative for data, outcomes and quality, which will position the

agency to collect and analyze existing data on behavioral health status, care delivery and patient, family and community outcomes throughout the U.S.

In addition to the work that's going on individually in the agencies, HHS convenes or participates in a number of quality work groups that address some aspects of behavioral health quality improvement or quality measurement, including the HHS Behavioral Health Coordinating Council, to advance HHS priorities in behavioral health, with a particular focus on integration with primary care.

The Interagency Working Group on Health Care Quality, which was established by the Affordable Care Act, to share information across agencies and ensure alignment and coordination between federal quality initiatives and the private sector; the HHS Quality Work Group, which is established to promote the development and implementation of the National Quality Strategy; the Measures

Application Partnership, which was established by NQF; and the HHS Measure Alignment Work Group, which is tasked with developing a process for reviewing and recommending measure alignment, new measure development and implementation, and measurement policy.

This is a timely and very important project and high priority for HHS, and this will feed into a number of our ongoing and planned activities. Some of our hopes for this project are the identification of relevant measure gaps in behavioral health performance measurement, in filling gaps in relevant measures and measure domains related to screening, assessment, follow-up and effective care, which can be used in all settings, and timely endorsement of additional ambulatory-based behavioral health measures for use in federal programs like Stage 3 meaningful use, for consideration across agency programs and activities, and for the development of a set of universal measures for consideration across HHS.

We hope that the measures considered here are able to maintain a focus on conditions that affect a large portion of the population, especially measures related to prevalent conditions among the ambulatory population seen in primary care settings.

So I hope that was at least some useful context for the discussion to follow, and we're very excited to see what you guys decide in the next two days. Thank you.

CO-CHAIR BRISS: Any questions or comments?

(No response.)

CO-CHAIR BRISS: Hearing none.

Project Introduction and Overview

MS. FRANKLIN: So at this time, we'll give a quick overview of our project. As Stephanie mentioned, this is a two-phase project, and for this first phase, we're looking at 21 measures that will be up for review by the Committee, and these primarily address tobacco and alcohol screening, medication management

and follow-up.

For Phase 2, we expect to be looking at approximately 48 measures, and these will address mental health conditions and maintenance measures. That number could change. We just are waiting to see what comes in in Phase 2. But this is the number that we've arrived at.

There might be a potential Phase 3 because of that large number that we're anticipating for Phase 2. So I just wanted to alert you.

So next, we'll go through for the Committee the measure evaluation review process, reserve status. That will be a very quick overview, and then any evidence exceptions and a related and competing measures discussion.

So as we're reviewing the measures today, we ask the Committee to go through the four major endorsement criteria, and these are -- and the hierarchy and rationale for each measure.

So to start off with, we'll describe the desirable characteristics of quality performance and measures for endorsement, and as we're walking through each measure, please call out and discuss the importance to measure and report for each measure, and that's looking at measuring those aspects with the greatest potential of driving improvement, and if the measure's not important on this criteria, other criteria -- I'm sorry.

This is a must-pass criteria that must be met as we walk through the measures. The second must-pass criteria is scientific acceptability of measure properties, and the goal here to ensure that the measure makes a valid conclusion about quality, and if not reliable and valid, the risk of an improper interpretation. This is also a must-pass criteria.

If the measure passes both importance and scientific acceptability, we'll be looking at the usability of the measure, and

you can see there the goal is to use the measure for decisions related to accountability improvement.

If it's not useful, you probably don't care if it's feasible. Then so feasible is the last piece of the picture that we'll be looking at, and ideally the measure should cause as little burden as possible.

And if the measure's determined to be not feasible, we could consider alternate approaches. If suitable for endorsement, and this is a yes-no question, we'll evaluate the measure in terms of other measures that are related, to see if there's harmonization needed and whether we will select a best in class measure.

So for all measures, both new and endorsed, they're expected to meet our current criteria, which has become more rigorous over the last several months, and for endorsed measures in particular, data from implementation of the measure as specified under

1(b), Opportunity for Improvement, should be looked at, and there's also here a potential for reserve status. We don't believe that there's anything in Phase 1 here that's going to fall into this category.

Reliability and validity testing should be expanded, unless it meets the right high rating per your review. When we get to the usability criteria, actual use in public reporting and other accountability improvement programs, or specific plans in a time line for use, is expected at this level.

We'll be looking at measures also for feasibility. If there's any problems with implementation or potential unintended consequences of that can be identified. Those should be identified here.

So here's our generic rating scale, and you can see here we have high, moderate, low and insufficient, and before you you have the definitions. We've been through this several times on the Committee. If we need a refresher,

you also have a quick -- you should have a quick guide in your packet that you can refer to throughout.

So the ratings we're looking at for high ratings should be based on information -- should be that based on the information submitted, there's a high confidence that certain data, that the criterion is well met.

A measure can qualify for a moderate rating here if it's based on the information submitted, there is moderate confidence or certainty that the criterion is met.

Low or insufficient, you can see the criterion there. Low, based on the information submitted, there's low confidence or certainty the criterion's met, and then insufficient, you simply find that there's insufficient evidence.

So to distinguish between a low rating versus a rating of insufficient evidence, a low rating generally means that evidence and information demonstrates that the criterion's not met, except quantity and quality of

evidence.

It depends on the combination of quantity, quality and consistency, and again we have that broken out for you in your quick guides.

Insufficient evidence means either the evidence does exist and was presented, but it's not adequate for a definitive answer, or the submission was incomplete or deficient in presenting evidence or information that does exist.

So if the Committee's, we rely on your expertise. If you're aware that there's evidence out there, we rely on you to call that to our attention. Ratings of low or insufficient evidence for subcriterion results, results in not meeting the criterion, but signifies different reasons.

So let's go on, let's move on to our importance to measure and report. So this is another -- this is our must-pass criterion, the first of two. It must meet all these three --

measures must meet all of these three subcriteria.

First is the 1(a) in the measure, high impact, which means you can look at the National Health Goal Priorities addressed, data on numbers of persons affected.

If there's high resource use or severity of illness or consequences of poor quality are high. You also look at 1(b), the performance gap, and we should be looking here for data showing a considerable variation in performance, or an overall less than optimal performance for the measure focus.

Data on disparities in care is really something we're looking at more and more, and that's very desirable in each of the measures. Potential for reserve status for endorsed measures, again, that's if there is no gap or a small gap. I don't think we have a measure that fits that description here.

Then we'll be looking at 1(c), evidence, and here we'd ask the Committee to

evaluate the quality -- quantity, quality and consistency of the entire body of evidence presented.

For subcriterion 1(b), performance gap, we'll be looking at variability in performance, overall poor performance, disparities in care, as mentioned earlier, and you should consider, as you're reviewing the measures, distribution of the performance scores, number and representativeness of the entities included in the measure performance data, and of course any data on disparities, as well as the size of the population at risk, and the effectiveness of the intervention.

Let's see. The reserve status, again, I don't think we have this. If you have questions, you can ask, we can discuss it offline, and we'll move on to submit it in existing evidence.

Individual Committee members will be looking at measures and rating them based on the evidence submitted. I know that if we're --

some members have said they're aware of additional evidence throughout the work group calls, and please call that to our attention as we walk through the measures again today, and continue to evaluate all the remaining criteria.

If we're confident that the evidence presented by the Committee members, when we walk through each of the work group discussions, and the measures likely to meet the criteria for high impact and scientific acceptability, we can then have a vote, discussion and vote on the measure.

There's also a possibility, as we walk through each of the measures, that we can ask the developers to make a change or provide additional information for the Committee, and reconsider any previous decisions on the measures.

And just another note about that, we are looking at the measure before us, and for the most part. Evidence rating scale. You also will have this in your quick guide, and you can see it there.

We're looking at the quantity of the body of evidence, and this is how we rate the number of studies. High is five plus studies; moderate is two to four studies; low is one study, and of course, insufficient to evaluate, no evidence or only selected studies from a larger body of evidence.

Next. Again, the quality of the body of evidence. We'll be looking at the certainty or confidence in the estimates of benefits and harms to the patients across the studies and the body of evidence. High, of course, is randomized control trials.

Direct evidence for specific measure focus, and there's an adequate sample size to obtain the precise estimate of effect, without any serious design flaws that introduce bias.

Of course, we definitely accept moderate evidence, which is non-randomized control trials with control for confounders, and if there's a large precise estimate of the effect

or RCTs without serious flaws. But either indirect evidence or imprecise estimate of effect.

A low rating would be RCTs with flaws that introduce bias, or non-RCTs with low, with small or imprecise estimates of the effect, or without a control of confounders, and insufficient to evaluate, no empirical evidence or only selected studies from a larger body of evidence.

So looking at the consistency of the results across the body of evidence, we're looking for both stability in the direction of magnitude of clinically practically meaningful benefits and harms to patients. A high rating would mean that the estimates of the clinically and practically meaningful benefits to the patient are consistent in direction and similar in magnitude across the preponderance of the studies.

A moderate rating would mean that an estimate of benefits and harms or in direction

that may differ in magnitude. For low, a low rating would mean that the estimates of benefits and harms offer both direction and magnitude, or wide confidence intervals that prevent estimating a net benefit.

If there's only one study, then the estimate of benefits do not greatly outweigh the harm. Of course, insufficient, no assessment of magnitude and direction or harms to the patients.

So subcriterion 1(c), evidence design logic. You can see it here, and I'll let you walk through this. I believe we also have this in our guide. This is a matrix showing how we arrive at our logic, for the quantity, quality and consistency of the measure.

So we'll go on to the next one. So we also have the exceptions to the evidence subcriterion, and I might ask Helen to weigh in if she has additional comments.

But if looking at the quality, quantity, quality and consistency of the

evidence, we can -- there's a question about passing on 1(c). We can invoke an exception, and this is in rare cases, for an outcome measure -- I don't think we have outcome measures in this phase.

But you can see for a health outcome measure, a rationale supports the relationship of the health outcome to at least one health care structure, process, intervention or service, and then it can be considered for an exception to the quantity, quality and consistency of evidence.

For our process measures, which fall into other types of measures category, if there's no empirical evidence except opinion, is systematically assessed with agreement, that the benefits to patients greatly outweigh the harms, we can look in exception, but only if there's consensus from the Committee that the potential benefits to patients clearly outweigh the potential harms. Otherwise, we cannot look at the exception. Helen, did you want to say

anything additional about that?

DR. BURSTIN: Just to mention again, it's intended to be an exception. We really want to rely on the evidence submitted in terms of quality, quantity and consistency. The other issue that comes up is at times, the submission may not have evidence that you're aware of.

You can certainly bring that to the table and we could ask the developers to add to that. But again, this is really intended for, as an example in our recent palliative care project, there was a measure specifically about the use of spiritual care and offering spiritual care.

Again, not something a huge amount of empiric evidence yet, but again one of those areas where the Committee invoked the exception, and said, for example, clear likelihood of benefits significantly outweighing risks in that area, and they put the measure forward.

But once we get through a few of

these, these things will be more obvious.

MS. FRANKLIN: Okay, thanks. Moving on. So as we just talked about, the exception. We draw the basis for our exceptions from the Evidence Task Force report guidance, and expert opinion is not empirical evidence, and will only be considered in exceptional circumstances, and I think Helen just ran through the conditions, so we can go on to the next one.

We'll move on to the next one. So moving on to the scientific acceptability of the measure properties, this is our second of the two must-pass criteria. We're looking at 2(a), reliability.

For each measure, you want to make sure that their specifications are precise, and that the reliability testing showing the -- sorry -- the reliability testing include the data elements of the measure score. For validity and threats to validity, we want to make sure the specs are consistent with the evidence,

and that the validity testing completed shows a significant data measure score, sorry.

For justification of exclusions, which relates to the evidence, the Committee should be satisfied that the justifications are valid for the exclusion criteria. We should also evaluate whether there's risk adjustment and whether risk adjustment would be a benefit or should be required for the measure.

Identification of differences in performance should be identified by the developer, and the comparability of data sources and method should be easily extractable for each of the measures.

For reliability and validity rating scales, again you'll have a shorthand for this at your place. You can see that for high rating of reliability, you want to see a precise specification and the empirical evidence of reliability of both the data elements and the measure score.

You want to make sure that the, for

validity purposes, the specs are consistent with the evidence presented, and the empirical evidence and validity of both data elements and measure score are valid, and threats to validity are empirically assessed and addressed.

Moderately rated would mean that the reliability of the measure specifications, they're precise, and the empirical evidence of reliability, there is empirical evidence of reliability, either the data elements or the measure score.

So again, here we have the low and insufficient. You're looking at the -- you would rate a measure low if you felt the specs were ambiguous, or if there was empirical evidence of unreliability. Looking at the validity, you make sure you rate it low if the specs were not consistent with the evidence, or there was empirical evidence of invalidity, or there were threats to the empirically assessed and biased results.

Insufficient, of course, there's an

inappropriate scope or method used in the reliability of the measure. In validity portions, there's inappropriate method or scope used and threats weren't assessed.

So evaluating the scientific acceptability, high ratings or moderate or high. You're going to have the yes, no questions here to see if, and then come to a consistence whether the measure passes scientific acceptability of measure properties for initial endorsement, and yes, if you have evidence of reliability and validity and no, of course, if there's inconsistent evidence of reliability, and reliability is usually considered necessary for a finding of validity, and you can see the rest of the scale there.

So once those two criteria are met, measures are past those two criterias, we can move onto usability in our discussions, and we'll be looking to the extent to which the intended audience will be able to understand and use the measure for decision-making. So it's

key that we're looking at the intended audience for each of these measures in this criterion.

So we're looking at 3(a), meaningful, understandable and useful for accountability in public reporting. We'll be looking at whether it's in use currently for public reporting or other accountability applications, or the measure developer has provided a plan, possibly a timeline.

Then there's also the possibility, the rationale=s provided. We want to make sure it's credible. In 3(b), we'll be looking for whether the measure is meaningful and understandable and useful for quality improvement, and is it currently in use for improvement, and if not, is there a plan to put it into use, and the rationale for using it for quality improvement is credible.

We are specifically looking at measures, though, that will be used for public reporting and not just for QI only. Okay. Feasibility. We'll be looking for whether

measures are readily, have data elements that are readily available and retrievable without an undue burden, and can be implemented for performance measurement.

Clinical data, under 4(a), clinical data generated and used during the pair process. For example, blood pressure lab values versus a survey or observation. 4(b), electronic sources, we're looking EHR extractability versus abstract and entering into a database or registry, and whether there's a credible or near-term path to electronic extraction or collection.

We'll also be looking at, under 4(c), the susceptibility to inaccurate or unintended consequences, in terms of the ability to audit and detect inaccuracies.

For 4(d), data collection strategy should be implementable, and are the data required by the elements already in operational use or is testing indicated that shows it's ready for operational use.

So that brings us to, once we've looked at all these four criteria, we'll be moving on to see whether a measure is related to other measures, and whether there needs to be a harmonization process that occurs, or whether the measure is superior to competing measures, or is more valid or efficient, or whether multiple measures are justified, so we can reach that conclusion as well. Helen, did you have anything additional to say?

DR. BURSTIN: We'll just get to it when we get to it.

MS. FRANKLIN: Okay. It's going to be a big part, all right. So let's move on. Again, we just talked about that, so we'll walk up through that in greater detail. That's related and competing measures later on. So let's see, what's our next --

CO-CHAIR BRISS: To quote Dr. Burstin, we'll get to it when we get to it.

MS. FRANKLIN: We'll get to it, and we'll get to that as well. So that's it.

MS. FANTA: Just to draw your attention to a few logistical issues, everyone should have a voting device, and let me know if you don't, as well as a folder. In the folder you'll find the agenda, the roster and that quick guide that Angela mentioned.

So as we're going through the measures, if you want to refer to that as you enter your ratings, that would probably be helpful. There's also voting instructions. The voting tool is fairly easy to use. We'll have the scale up as you vote.

You basically just need to press the number and make sure you point to Evan when you do so, because this is where it's gathering all the signals. If you want to change your vote, you have a minute to do so, or we can always just restart the voting period.

But if you want to revote, you can just press the button that has the little exclamation point and then re-enter the number. No need to press send or anything. And then

there's also in the folder a logistical form for any reimbursement issues that you may have as you are in D.C. for the next two days.

Does anyone not have a voting device or a folder? Make sure we get you in. Okay, two. Okay, thank you.

CO-CHAIR BRISS: And then so on the NQF jurisprudence issues that we just went through, I sort of -- I would tend to agree with Helen, that this will be easier to talk through as we go through specific examples.

But Dr. Pincus has a question, and then I think we have a couple of minutes for a general discussion, if we need to do that.

CO-CHAIR PINCUS: So in -- I bring this up often. Helen gets annoyed when I do, but there's always the issue of whether we're rating the measure concept versus the actual measure, and measure concept being the concept to which it's being applied, rather than the measure as specified with those intended denominators, numerators, processes, and also for those

applications, as described by the measure developer.

It becomes particularly relevant under two issues, and some clarification so we get some consistency, I think, would be helpful. So one comes up in terms of rating the body of evidence. Are we rating the body of evidence for the concept or for the actual measure?

And particularly in terms of thinking about, you know, counting studies. Is the study looking at the utility of screening and brief intervention for tobacco use, for example? Is it all studies looking at that issue versus the specific measure that's being proposed, and studies utilizing just that specific measure?

It also comes up in terms of the notion of validity, to some degree, that when -- and in terms of rating of evidence about validity.

A number of the measures that we're dealing with have, use a convenient sample of the survey for face validity. To what extent is

that a measure of validity versus actual measuring the association of the processes being measured with particular outcomes?

If it's measuring that process, is it actually using that measure to measure the process, and is there a study documenting that? So it would be helpful to have some guidance on that, and I think the third point is a number of these measures are a suite of associated measures, and the question of, as a suite, evaluating them versus as individual measures.

In some cases, there's a measure that's established really as an anchor point, that in and of itself probably would not be a great measure, but it's essential for measuring the suite of measures. And just to get some guidance about sort of dealing with those three issues.

DR. BURSTIN: Those are good issues. We've talked about these before. So some of you may have heard that NQF is moving to a new process, but probably not really beginning

pilot testing some time later this year, beginning it in early 2013, of likely moving towards a two-stage endorsement process, with the first stage actually looking at the measure concept, really mainly around importance to measure and report, and then having the developers come back when they're ready to do the second stage, which should be the rest of the criteria on a fully-tested, fully-specified measure.

Part of the logic for that is we get a fair number of measures in that probably aren't quite ready in terms of the second half, but are very ready for an early read in a sense of the evidence and a sense of the importance.

That's not where we are today. We are still looking at measures that are coming before you, fully specified, fully tested. So you do need to evaluate the measure, not the concept of what this measure could be. We can't do measure development on the fly today, much as many of us with development expertise sometimes

get very tempted to do.

There are at times, it's reasonable at times to raise some very specific issues. For example, one specific exception that you think, you know, significantly throws off the validity of the measure.

Those are issues that the developer is able to consider, come back, indicate whether they may perhaps be willing to take the advice of the Committee. But they need to be fairly small issues, certainly not anything grand.

In terms of evidence, you know, is it evidence for the measure area in general or the measure specifically? The way the criterion is written, it's specifically evidence for the measure focus. So it is about the measure.

But we truly understand that in some of these areas, we are oftentimes invoking evidence from a slightly broader viewpoint than the specifics of the measure, and that's where your expert opinion comes into being, of whether

you think that's sufficient or not.

In general, we'd like the evidence for the measure focus be as specific to the precise way this measure is in fact worded, but we recognize at times that's not always going to be available.

In terms of the issue around validity, we do recognize that face validity certainly is not the highest level of validity, by any means, but we do consider it the floor.

So if a measure at least has a systematic assessment of face validity, that is acceptable, but again is rated the lowest acceptable level for face validity.

The last issue raised about individual measures versus suites of measures, we only endorse individual measures. We don't endorse a set of measures or anything along those lines. The one thing we do do is we do in fact endorse pairs or composites.

So some of the measures you've mentioned, I think, referring to Harold, are

measures that are really anchors for the other measure.

Those are paired measures, and we can determine as we get to those if that other measure doesn't make it, and the one to which it's anchored is really just about adjustment or something like that for the initial measure, we can decide those as we come forward.

But at this point, we don't endorse sets of measures or suites of measures, and we recognize that that can be difficult for the developers, who in fact have created a suite. But if one or two of them don't pass the criteria, then we can move them forward.

CO-CHAIR BRISS: And having said that, having said that in terms of the efficiency of our discussion today.

DR. BURSTIN: Yes.

CO-CHAIR BRISS: We're going to do two sets of four measures that are related to each other. So in terms of chairing the meeting and in terms of trying to be efficient with the

use of our time, so assuming, for example, that we establish hypothetically that tobacco's a really important public health problem, we presumably won't have to readjudicate that six times today.

So we may be able to -- hopefully we can, when we're evaluating suites of related measures, we may try to have the discussion the first time, and perhaps be efficient about repeating in subsequent related measures.

CO-CHAIR PINCUS: Although I think we do have to formally vote on each one.

DR. BURSTIN: Yes, right, and there may still be nuances. I mean smoking may be a high impact area, but keep in mind there are still two other subcriteria and importance to measure and report, and in fact the performance gap may be very, very different from an assessment measure versus a treatment measure, and even the evidence underlying, in fact, the intervention may be very different than screening.

So I think it's more about, sure, you can take a pass on high impact, but you've still got other stuff to do. Was there a question back here, Tami?

DR. MARK: So as we try to assess and then vote on these measures, in terms of the formal and relatively empirical criteria that you laid out, as part of the process, are you going to review, you know, that criteria so, you know, someone will say there are five RCTs on this and the kappa statistic is this, or are we supposed to bring that to this meeting?

DR. BURSTIN: That's part of the reason for you having your quick guide, which is what Sarah was just mentioning. It's actually in your folder.

So it's a very simple, just somebody tried to tell me to make a two-pager. That was impossible. It's a four-pager, but it's on two pieces of paper, is my justification.

But it includes, for example, all those tables that will allow you to make that

assessment. But we'll be happy to weigh in --

DR. MARK: I'm not sure my question was clear.

DR. BURSTIN: Okay.

DR. MARK: For each particular measure, are we going to have a summary of how many RCTs exist and what the kappa statistic was for that particular measure?

DR. BURSTIN: You'll have whatever was submitted on the form, which we'll be projecting, yes, and the people who have been the primary reviewer will obviously take a deeper dive on those issues.

DR. MARK: Okay, because the form -- okay. So we'll just project the form up.

DR. SUSMAN: So I have a question that mostly is around validity. In thinking about measures that are for performance improvement, things like risk stratification probably aren't horribly important, because there's an internal effort to improve care.

But risk stratification and a number

of the topics under validity often don't have any demonstrated evidence, in my experience, with this, and these measures seem to be no exception. How much do you see us weighing that, particularly when we're looking at measures that might be for public accountability?

DR. BURSTIN: Those are the kind of things that you're going to have to discuss as a group. I mean we give you as much guidance as we can there, but those are assessments you'll need to make.

Can this measure stand as a consensus standard that could potentially be used for public reporting, pay for performance, a whole variety of potential applications as is, or does there need to be additional work to kind of level the playing field?

Those are exactly the kind of issues that would of course come up under some of our discussions around validity and risk adjustment, and stratification as well.

CO-CHAIR BRISS: Anybody else have

questions or comments or concerns about the jurisprudence?

DR. SCHMALTZ: I was just wondering, in terms of as we review each measure, will what was discussed on the conference calls be shared with the group?

CO-CHAIR BRISS: Yes.

DR. BURSTIN: And I actually wonder if we want to actually distribute that paper.

(Off-mic comment.)

DR. BURSTIN: I think we may not -- I'm just not sure everybody's going to be looking at SharePoint and their thumb drive all at the same time. So maybe I'll just ask my assistant to print out the summary of what you guys did on the work groups. We find that tends to be helpful.

CO-CHAIR BRISS: Anybody else? Questions or comments or concerns?

DR. BURSTIN: The first measure usually takes 90 minutes. Don't worry about it. It's okay. We'll catch up, and it's kind of a

learning process. Go ahead.

DR. HANRAHAN: Can you give me a sense of what the longitudinal plan is for these measures, in that I know there's a process where they go back in to be re-reviewed at some point, because there are measures that we've looked at that really have low scientific quality, but their usability and feasibility are pretty good.

So you know, I'm just wondering, if we do approve a measure that has some value to moving along the agenda of quality, what's the pathway that it's going to take over time, and can we recommend some of the science that needs to happen with that measure, to really improve the quality of the measure?

DR. BURSTIN: It's an excellent question. So all of our measures routinely get evaluated every three years, as a matter of course. That's when maintenance comes up. The three years was chosen very much to be along the lines of guidelines in the evidence base. Typically, that's around when you do typically

see changes in evidence.

We do also have a process of what's called ad hoc reviews. At any point at any time anyone can come forward and say there's a problem with this measure. Either the evidence has changed or there's an unintended consequence out there. So for example, several years ago there was a measure out there about pneumonia in emergency departments, you know, trying to get antibiotics in within four hours, and it was very clear that a lot of little old ladies with CHF were getting antibiotics for pneumonia, not the intent of the measure.

So that was recognized. We did a re-review of that measure. That measure was modified to make it clear, presumptive diagnosis of pneumonia, and you know, a series of changes. So at any point, implementation concerns could be brought forward to trigger an ad hoc review.

Expansion of settings, for example, oftentimes triggers an ad hoc review. This was a measure previously used only in hospitals,

they'd be great in nursing homes. We can evaluate it that way. The third one is really a change in evidence.

So as that science emerges and something comes forward, at any point we can also reevaluate that measure. All developers are also required to do an annual update to NQF, not so much in terms of the evidence, but literally, here's the specifications, has anything changed? If so, why has it changed?

And any of those annual updates if they're significantly different can also trigger a real ad hoc review for us as well.

CO-CHAIR BRISS: Was that all of your question, or were you also asking what happens to these measures after we finish with them over the next couple of days?

(Off-mic comment.)

CO-CHAIR BRISS: So anybody else? Okay. So are we ready for the first measure? So yes. So I'm about to take off my chair's hat and become a Committee member. I made a serious

tactical error by agreeing to be first.

MS. FRANKLIN: So that brings us to Measure No. 1651, and we would ask that for each measure, the developer present the measure, give us a quick overview, and then we'll have the lead discussant start the discussion of the measure.

This is the Joint Commission measure 1651, TOB-1, Tobacco Use Screening.

Measure 1651

MS. LAWLER: This measure looks at screening -- can you hear me now? Okay. This particular measure looks at screening all hospitalized patients age 18 years of age and older for tobacco use, and that's all tobacco products, be they cigarettes, cigar, pipe, smokeless tobacco products, and the denominator is all patients.

So it's kind of a global measure. Again, 18 years of age and older. It's the adult population, and the numerator is the number of patients who were screened for tobacco use status. So again, looking at all patients and

then the number that were screened for the tobacco use status.

The exclusions are, of course, just those that are less than 18. Patients who can't answer a question, so they're cognitively impaired, and you're just not able to screen them. We did have a length of stay exclusion here for patients who are there for an exceptionally long period of time.

This has to do with our specifications, with an alignment with CMS, and working over manuals. So sometimes there's some problems with that.

MS. WATT: This is Ann from the Joint Commission, and just as a little background, this is a relatively new measure set for us. It has just gone into use for purposes of Joint Commission accreditation, just beginning January 1st of 2012.

We tested the measures in 2010, over a six month period of time. So all our experience is relatively new. However, they

are available for hospitals to choose as part of their ORYX accreditation requirement.

I think it's also important to point out that this measure set was developed for use with all patients, not just behavioral health patients, and in fact, we had thought that it as going to be discussed under the Preventive Health Project. So I guess that's it.

DR. FIORI: And just very briefly to present the science, the scientific rationale for this measure, there's substantial data that when tobacco users are identified, you substantially and significantly statistically increase the likelihood that clinicians will then go on to provide evidence-based interventions.

There's also some evidence that solely by screening, you actually have an impact on downstream quit rates. So it not only is a central measure for identifying who should be a target for tobacco cessation interventions, but it also, just by virtue of the screening, has an

impact on both the rate at which clinicians deliver cessation interventions and downstream quit rates.

CO-CHAIR BRISS: So as the work group has sort of worked through this measure, so shockingly enough, everybody rated the leading preventable cause of death in the United States, a potentially high importance condition, right.

So that the data on performance gap ranged from about 60 percent in the reviews of existing studies, and was more like 70 to as high as 90 percent in the piloting of the measures. So people that thought in general, that performance gap data was either high or moderate.

The rationale or the scientific evidence that Dr. Fiori just gave was generally rated also either high or moderate. So good evidence that screening provokes effective cessation treatments and probably moderate evidence. I think it was three studies on effects on cessation itself.

In terms of reliability and oh, I'm sorry, and the quality and consistency of the body of evidence was generally considered to be high or high to moderate.

So overall, in terms of the importance to measure and report criterion, the work group voted 6 to 0 that it passed this criterion, and with that, I'll stop and take a breath and we can have a discussion about that section.

CO-CHAIR PINCUS: So what would be useful is maybe to go through each of the criteria on which we have to vote, and then sort of just -- sort of hit that issue, and then see if there's any discussion on that issue, okay. So for the first criterion on which we would have to vote is --

MS. FANTA: So that's 1(a)?

CO-CHAIR PINCUS: --importance. So Peter, is there anything more that you want to add on that?

CO-CHAIR BRISS: I'm sorry. Would anybody else from the work group like to add to

the summary?

CO-CHAIR PINCUS: Now is there any discussion, any questions that anybody else on the Committee has?

(No response.)

CO-CHAIR PINCUS: So I guess we're prepared to vote.

CO-CHAIR BRISS: Wait.

CO-CHAIR PINCUS: Oh wait.

(Off record comments.)

CO-CHAIR BRISS: Yeah, 1(a).

CO-CHAIR PINCUS: So I guess we vote?

MS. FANTA: We can go ahead and vote on 1(a), so go ahead and point over it.

CO-CHAIR PINCUS: If we press more than once, what happens?

DR. BURSTIN: Nothing. You only get it once.

CO-CHAIR BRISS: So the numbers are -- we push the number corresponding with that?

MS. FANTA: So it's 1 high, 2 moderate, 3 low, and 4 insufficient.

CO-CHAIR PINCUS: Point at Sarah.

MS. FANTA: Yes. So right now, we have 15 people who have voted, so we're waiting on four more. So if everyone could just point again over at me. Now we're at 16. No, it will not count you twice, to just keep voting.

(Pause.)

MS. FANTA: So we have 19 high.

CO-CHAIR PINCUS: Peter, any further comments on the performance gap?

CO-CHAIR BRISS: So again, the data that was reviewed went from as low as 60 percent performance, based on the literature review, to more like 70 to 90 in the piloting.

CO-CHAIR PINCUS: Any further comments by the work group?

(No response.)

CO-CHAIR PINCUS: Any questions or comments by the overall Committee?

DR. SAMET: I'm just curious. Did it matter if the site, or was there any discussion of the site where the question was posed in the

hospital setting, as opposed to where evidence, other evidence lies?

CO-CHAIR PINCUS: Are you saying is it sort of the evidence for being it in a hospital, doing the screening in a hospital, versus screening in a primary care clinic?

DR. SAMET: Yes, right, right. We're specifically asking about hospital here, and yet a lot of data is in other places.

CO-CHAIR BRISS: I think that the developers may want to comment about this. My recollection was that all of those data came from hospitals. I don't know if --

MS. LAWLER: Just to clarify the question, is the evidence specific to inpatient settings versus all settings? Is that what the question is, on the performance gap? Dr. Fiori, do you know?

DR. FIORI: I'm sorry. So is the 70 percent number that you mentioned, is that what the question is relating to that? I believe that came out of the testing hospitals.

CO-CHAIR BRISS: Yeah. So there were two kinds of -- essentially in my synthesis, there were two kinds of data that you reviewed in the application, that bore on the performance gap. So there was about a 60-ish percent number from the review of literature, and then a more like 70 to 90 in the testing hospitals.

DR. FIORI: Correct.

CO-CHAIR BRISS: So the question was, was all of that from hospital settings and not other settings, and what do you think, and perhaps what do you think about the generalizability of those data?

MS. LAWLER: The answer to the question is that the statistics from the testing were from hospitals only.

DR. MARK: Hi. Looking at some of the literature, it describes the reviews as interventions, and I'm wondering if they're focused more on treating tobacco, in terms of how you define the gap, or is it really screening?

So you're saying there's a 70 percent

gap in screening for tobacco? Hospitals aren't asking patients do they smoke, or is the gap about the intervention that happens after they screen?

CO-CHAIR BRISS: For this particular measure.

DR. FIORI: In this particular measure, it actually related specifically to screening, and the literature cited was specifically to screening. There's additional literature that we'll talk about later regarding the intervention points that you just made, ma'am.

CO-CHAIR BRISS: And as Helen said in the run-up, we're going to be reviewing a suite of related measures, that talk about -- that essentially talking about screening first and then intervention and then follow-up, and the gaps that people talk about will be related to those specific issues.

DR. PINDOLIA: When I was reading the literature and the stats, and when you said it,

it still troubled me, because I can't figure this part out.

So in the literature, it's 60 percent or so, but in the 30 hospitals it was close to 70 to 90 percent that were getting the screening, and they were using the ORYX data collection in the 30 hospitals that JCAHO provides?

MS. WATT: Yes. We had a special data collection tool that they used. They could use their own screening tool, but they put data into our data collection tool. So yes.

DR. PINDOLIA: So the question I have, I know for other quality measures and tools that we have, when there's a tool developed specifically to collect that data, even though the work was already done, it's not collected when you don't have the right tool. Like in EMR, they might not have a specific place to have the patient was screen for tobacco screening inpatient.

But when you get a tool in there, you'll see that the rate goes higher. So I'm

trying to figure out is the gap really 60 percent, or is the gap closer to 70 to 90 percent.

It's just that they didn't have the right tool to capture it in the literature, when they did that analysis. Am I making myself clear? I'm trying to figure out if the gap really is as low as 60 percent?

Or is it really closer to the 80 percent average that the 30 hospitals that had a tool to enter data, so you can actually specifically capture that point, whereas in the studies and that, you go through and you do data collection and you don't have a specific tool to collect it.

So maybe the gap isn't as bad, and that's what I'm trying to understand.

CO-CHAIR BRISS: Isn't it true that your tool was used retrospectively in this context? You were going back and it wasn't that people were using the tool at the time of patient care; it was you were using the tool retrospectively after the episode of care; is

that right?

MS. WATT: That's correct, yes.

DR. PINDOLIA: Thank you.

CO-CHAIR PINCUS: Any other comments?

(No response.)

CO-CHAIR PINCUS: Okay. So we're ready to vote on the performance gap. So everybody point at Sarah and vote.

MS. FANTA: So we have 11 high and 8 moderate.

CO-CHAIR BRISS: So in terms of quantity of evidence, the data that were presented, and again by Dr. Fiori this morning, there are lots of studies suggesting that screening improves the delivery of cessation services. There are three studies, I think, that suggest that screening increases rates of cessation.

CO-CHAIR PINCUS: Any comments, questions, by any members of the Committee?

DR. SUSMAN: So my question specifically here is whether the intervention

effects by screening are attributable to inpatient hospitalized settings, or is it a generalization from outpatient-oriented settings, where screening has a fairly clear link to reduction in rates of smoking?

DR. FIORI: The data from the meta-analyses that were part of the Public Health Service Clinical Practice guideline included both types. So you're correct. It's not exclusively to hospitalized settings.

We have no reason to think that it would be different in hospitalized settings in a substantial way, but you are correct. The data was for all settings.

CO-CHAIR BRISS: There might be a conceptual reason to think that, sort of based on the burned hand teaching best, that a hospitalized person might be in a teachable moment, and might be at least as likely as the person on the outside.

DR. SUSMAN: I think you could probably argue it both ways. If you're there

for a heart attack, why yeah. If you're there for your maybe orthopedic procedure or some unrelated thing, then perhaps not.

DR. MARK: I'm looking at the criteria. Can you say whether the studies were RCTs or not, and also address this issue of consistency.

CO-CHAIR PINCUS: I know. This is all together.

CO-CHAIR BRISS: Oh, it's all -- I'm sorry. It's all together.

CO-CHAIR PINCUS: Yeah. Quantity, quality and consistency.

CO-CHAIR BRISS: So I think these were all trials, right?

DR. FIORI: Yes. They were all RCTs.

CO-CHAIR BRISS: And --

DR. FIORI: With other criteria, they would have at least six months of data. So there was a rigorous *a priori* set of criteria to be entered in the meta-analyses, which included RCTs.

CO-CHAIR BRISS: And can you comment about the consistency issue?

CO-CHAIR PINCUS: Just one other question around that. When you say it was an RCT of just screening, nothing else associated with it?

DR. FIORI: The question, it would sometimes -- so the simple answer is yes. It sometimes was part of a study that actually also looked at some other outcome measures, but that was independently looked at as a question.

CO-CHAIR PINCUS: No, but the question I have was the intervention just screening, leaving aside the outcomes? In terms of the meta-analysis of RCTs, it was just screening? No counseling necessarily associated --

DR. FIORI: Right. You would have to -- otherwise, it would just be confounded in a way that you couldn't tease out the screening part. So yes.

CO-CHAIR BRISS: And the consistency

question? You want to -- can you comment on the consistency of this body of evidence?

DR. FIORI: You mentioned that there are two outcomes that we'd looked at. One is does it increase the rate at which clinicians provide smoking cessation interventions that are evidence-based?

That is highly consistent, and there's a large number of studies. To the second question, does screening by itself result in downstream increases in quit rates, that data, while suggesting it does, was less consistent in a smaller number of studies.

(Pause.)

DR. FIORI: Correct, 3.1.

DR. SAMET: So just to get clarity on this issue of site in which it's being carried out, maybe the way to ask the question is was there any difference in the evidence when these studies were done in the inpatient setting, than it was done in the outpatient setting? Because you said they combined some of them. So I'm just

-- did they look at that?

DR. FIORI: The results were consistent across all the studies we looked at. Dr. Samet, I don't have in front of me all of them to be able and go through and give you that specifically. But it was consistent across all of the sites.

CO-CHAIR PINCUS: This case, there is a different scale?

MS. FANTA: Yes. So please press 1 if the evidence is sufficient, 2 if it's not or 3 if it's insufficient evidence.

(Off record comments.)

CO-CHAIR PINCUS: Peter, is on reliability. We're looking at the acceptability of the -- the scientific acceptability of the measure properties, virtual reliability and then validity.

DR. BURSTIN: We passed it.

CO-CHAIR BRISS: We've passed importance to measure and reported, and we're moving on to the second one.

DR. BURSTIN: So scientific acceptability is three subcriteria, all required to be passed. All three were passed, so you can move on.

CO-CHAIR PINCUS: So we don't need to go to reliability and validity.

DR. BURSTIN: No, no.

(Simultaneous speaking.)

CO-CHAIR PINCUS: Okay. That's the next thing.

CO-CHAIR BRISS: So in terms of reliability and acceptability, you may want to move forward to that point in the -- so generally, yes. So generally they reported, the text suggests reasonable agreement among coders, in terms of reliability.

It was interesting that the text in the kappa statistic looked to me to be discordant. So the kappa that was reported was actually low. It was .03 or .05 or something along those lines. So I'd love a comment from the measure developer about reconciling the text and the kappa.

Overall based on the text, the subgroup generally scored reliability as high or moderate.

MS. WATT: With regard to this measure, the kappa score was low, and sadly, it was not a misprint. I think what the problem was in discussing this with our statistician, the reason for the low kappa was because the disagreement on the number of denominator cases during the testing, and it was --

The reason why we test measures is to figure if the specifications that we write are clear and understandable, and in this case, we found no -- they weren't clear and understandable, because we were getting different results than the test hospitals were.

So what we did as a result of this and went back and made revisions to the measure specifications, in order to correct the problems that we identified during the testing process. So that's the reason.

CO-CHAIR PINCUS: Do you know --

CO-CHAIR BRISS: I'm sorry, and what were the issues on the denominator on which they were having disagreements?

MS. WATT: It was twofold here, as we had discussed in the submission here. The wording in the data element allowable values that made people confused about whether or not someone refused a screen, or the screen wasn't offered.

So it was just tweaking of that wording that needed to be more clarified, and then conflicting information that one might find in the medical record, about whether or not the patient is using tobacco products or how much they're using.

And so to correct that, we made several changes in our specifications for that data element, giving the abstracters more clear guidance on what to do if there is conflicting information. So we feel like the changes that we made are going to really make a big difference in how reliable the data are.

CO-CHAIR BRISS: There are a lot of people stacked up on this question. So what you just said was a denominator problem, but that sounds like a numerator problem as well, right? So if you were having trouble assessing whether there was screening, that's a problem too, right?

MS. WATT: Right, right.

CO-CHAIR PINCUS: Just to finish off this question, was any further testing done to see if the new, improved methodology resulted in any change in reliability?

MS. WATT: Just to clarify, it's not a change in methodology. What we have done is strengthen the specifications, and no, we have not -- well, the testing now or the retest now is the actual experience, data collection experience, which began on January 1st, and we will assess that.

We assess our measures on a semi-annual basis every six months, and address necessary revisions to the specifications and

that kind of a thing.

So this will be addressed. We also have, and I don't want to spend a lot of your time, but as you may know, we have a network of contracted performance measurement systems or vendors who collect the data from hospitals on behalf of the Joint Commission, and they are required contractually to do continuing reliability studies, and we get those data from them as well.

CO-CHAIR PINCUS: So Jeffrey, you have a question?

DR. SUSMAN: Well, it's really a follow-up on that. I guess I'm a little bit concerned that the kappa was so low, and that we don't have the data that validates or looks at that in the revised measure specification. So that's the issue that's out there for me, at least.

CO-CHAIR PINCUS: Any other comments or questions about the issue of reliability?

DR. BURSTIN: I just have a question

for Ann. Ann, so will there an analysis on further kappa statistics that could be provided, and if so, when?

MS. WATT: Yes, there will be and yes, they could be provided, and depending -- and actually we'll have to consult with our statisticians on this one. Generally speaking, the data come to us on a quarterly basis.

Assuming that that's a sufficient amount of data, we get that four months after the close of the quarter. It takes time for analysis. So we're talking essentially end of 2012.

CO-CHAIR PINCUS: Okay, Jeff.

DR. SUSMAN: So maybe this is just a Helen question. In the new NQF process or what we're operating under, could we potentially ask a measure developer to hey, bring us back the further data and then resubmit, or does this mean we've deep-sized it for a long time? What are the outcomes?

DR. BURSTIN: Yes. It's still not

clear exactly when we would do this set of measures again. I think one of the things we'll need to think about is just looking at the other data presented broadly, in terms of reliability and validity. Is this a no move forward for you, or is this something that you think, you know, there was a potential --

We could of course have the Joint Commission forward that as part of their annual update next year, if the measure gets forwarded. But again, that is something you need to weigh in. That is, you know, obviously quite a low kappa.

DR. FIORI: This is, hopefully will be helpful information, but screening is the more straightforward measure. So identifying tobacco use or not tends to be, apart from these few hospitals where it was tested, these 30 hospitals, in general it tends to have high reliability.

That's supported by some evolution of the electronic health record, which is in a more

and more consistent way asking about tobacco use. So that's an additional piece that I think will help reliability enormously when this is in the field.

And then finally, through meaningful use and some other measures, the way that tobacco use is asked is becoming standardized, and that wasn't the case when we were in the field here.

But I think it will be in the future, and that will help substantially to increase reliability, particularly on whether a person does or doesn't use tobacco.

CO-CHAIR BRISS: So let's move to the testing results. These were the testing results that were reported, so and this was a bit confusing in the subgroup.

So they essentially reabstracted 130-ish cases. Only a couple percent of those had a false positive, and only a few percent of those had a false negative.

So based on this, the Committee generally thought that that seemed like

reasonable performance, and truth is, we thought at that point that the kappa statistic might be a misprint. So that's the rest of the universe of information that you might consider, as we're thinking about whether this is sufficiently reliable to allow to go.

CO-CHAIR PINCUS: Jeff Samet.

DR. SAMET: So, you know, under these moderate rating, it says systematic assessment of face validity. I was just trying to -- maybe this fits within that, what you just said. But I wasn't quite sure if it does. So I was looking for some clarity of what that mean, this systematic --

I mean, it has face validity to me, but maybe by seeing it there, that's systematic. I don't know. What do you think?

CO-CHAIR PINCUS: Well, I think, just to clarify, the reliability issue, you know, is looking at the kappa and, you know, a kappa of .031 is no better than chance. Tami?

DR. MARK: I'm trying to think through

this issue. I'd be interested in getting people's opinion on what the negative consequences might be of implementing a measure that was not reliable, and you know, and do you think that would lead to, you know, that would be a short-term thing?

So hospitals would realize the measure wasn't reliable. They'd quickly improve it. There would be no negative consequences, or is that something that you think would be particularly bad?

CO-CHAIR PINCUS: Peter, do you want to respond to that?

CO-CHAIR BRISS: Yeah. I mean I can give -- so it depends, right. So if you ask an epidemiologist a question, the answer is always Ait depends,@ right? And then they say that you're well-trained, you can talk about why it depends.

So if it's for quality improvement, it depends on the use to -- it seems to me it would depend on the use to which the measure might be

put. So if you're using this for internal quality improvement, it strikes me that it might not be so terrible if there were some fuzz in the measure, particularly if it was consistent within a hospital.

If you were using it for a high stakes use of something like comparing hospitals, it might matter a whole lot if there were fuzz in the measure, especially if it were inconsistent across hospitals and created a systematic bias that actually made a hospital look better or worse than it really was.

CO-CHAIR PINCUS: And just to point out, that we're evaluating these on the basis of both use for improvement as well as for accountability. It's got to be both.

CO-CHAIR BRISS: Anybody else have thoughts about consequences?

CO-CHAIR PINCUS: Vanita.

DR. PINDOLIA: Hi. I'm sorry. I'm still going back to is there really a gap, because as the Joint Commission said, that now,

with more EMRs, and I know with our system transferring to Epic, and I think, what, 30 percent of the nation's population is there, it's like a mandatory question. You have no choice but to ask that.

So is this just a natural evolution that's this is going to just be there, and it sounded like Joint Commission is saying that that's what they're seeing with the EMRs. It's transcending down that line.

I'm trying to understand what is the need -- is there a need for this measure at this point?

MS. WATT: Excuse me, just one second. I'm corresponding with our statistician here. He's on the line and actually would like to talk, but the operator won't patch him through. Is there any way that we can --

OPERATOR: Yes. If you would like to open the floor for questions from the phone, they may press *1 to have their line opened.

MS. WATT: Steven, dial *1 please.

Thank you.

OPERATOR: And Steven Schmaltz's line is open.

DR. SCHMALTZ: Hello, thanks. I just wanted to make a comment about the kappa. The kappa is a chance-corrected agreement statistic. So if most of the cases fall in one particular category, the kappa will really be sensitive to disagreement in the lower proportion category.

That's what happened here. 94 percent of the 131 cases were actually a Category E. There were very few Category Ds, and those few Category Ds are the ones that have the --

CO-CHAIR BRISS: Can you translate Categories E and D for us please?

DR. SCHMALTZ: Category E means they're compliant with the measure. So that means -- this is a screening measure. That means they were screened, and then a Category D means they fall into the measure, but they were not screened. So the disagreements were among

those that were not screened, rather than those that were screened.

CO-CHAIR BRISS: So that might make me feel better, actually. So there was then perhaps, then perhaps in -- let me test this out. So there was high agreement in general about the numerator of the measure, and relatively high agreement about the denominator of the measure, and because so many of the cases fell into the compliant people, the little bit of the fuzz in the denominator, it sort of made your kappa score be very low. Is that fair?

DR. SCHMALTZ: That's correct.

CO-CHAIR PINCUS: All right. But just let me interject. But then that -- what you're saying, though, goes against the fact, the finding that there's a gap. If you're saying that in the pilot, that there was virtually universal screening, then that raises a question about whether there's a gap.

MS. WATT: Excuse me. The pilot looked at agreement rates. So it doesn't say

that 90-some percent met the measure. It says that 90-some percent of the time, the Joint Commission re-abstracters agreed with the abstraction of the original abstracters.

We weren't looking at measure rates themselves as part of that reliability study. We did compute overall measure rate for the pilot, but that's not what that 90 percent represents.

CO-CHAIR PINCUS: Then that needs further clarification.

DR. SCHMALTZ: It turned out the hospitals that were in the reliability study had high rates on this measure, but that doesn't imply that that high rate is appropriate across the whole population.

CO-CHAIR PINCUS: I agree. That's my -- I guess my point is that it raises a question about if this was more broadly applied, how generalizable is the reliability study.

CO-CHAIR BRISS: Harold, so I think we might benefit from separating a couple of

issues. So we had already -- I thought -- Helen, this is a jurisprudence question -- I thought that we were supposed to make judgments based on the information that's in the application; is that correct?

And so we can speculate about the time trends on -- I would have said based on that, that us speculating about the time trends of how this measure might be, get to have no additional gap based on the further adoption of electronic medical records. But I would have called that out of bounds for us, in terms of whether there's a remaining gap.

DR. BURSTIN: I agree. I think just the other thing I want to point out though is it's typical that there's in fact oftentimes a very specific population who agrees to be in a pilot. So it may not be generalizable. I think that the rate, which is extraordinarily high for hospitals, 99.6 percent, I think, again, those were self-selected hospitals that agreed to be in a pilot.

We don't know what the number would be. The EHR issue is an interesting one, because I think it does point out just broadly the question of whether some of the low rates of what you're finding is in fact the lack of documentation in a paper chart, as opposed to if you did this in an EHR, whether it would of course be universal because it's always there.

At this point in time, you're evaluating the measure as a paper measure, which is how it is put forward. Hopefully over time, they'll move to making this a measure off EHRs, which I think will be a far more effective measure.

CO-CHAIR BRISS: So then, and in terms of the reliability, in terms of the reliability, it seems to me that we're back to the data that's here, which is generally good agreement on a denominator -- I'm sorry, generally good agreement about on the numerator; somewhat less agreement on the denominator, and a low kappa because of the second point.

And so we might -- does anybody have additional questions or comments?

DR. SUSMAN: The only other thing that I'd like to point out is that this often will be used as a paired or a sequenced measure, that are going down a path of identification, intervention and follow-up.

To me, that probably weighs in the favor of it, because it's really the basis for being able to do anything substantive around intervention, and being able to, you know, link this ultimately to some further downstream outcomes.

CO-CHAIR BRISS: And it is true that the gaps in performance get bigger -- as we go through the rest of these measures, we'll find that the gaps in performance generally get bigger.

MR. WILLIAMSON: So we're ready to vote. All right. The timing is now started. We have a minute. It looks like we're doing a little better this time. One more. All right.

So for the reliability, we have 3 high, 7 moderate, 6 low and 3 insufficient evidence.

DR. BURSTIN: It's actually right on the bubble. Actually, could you go back for a second?

Right. So it's 10 to 9, essentially. We require high or moderate to move forward. I guess we'll proceed, but it is certainly not consensus.

CO-CHAIR PINCUS: Let's move on to validity.

CO-CHAIR BRISS: Sorry. So essentially for face validity, the short answer is they asked a lot of experts and hospitals or some set of experts and hospitals to assess face validity and it generally scored high, and if you can page forward to where those scores show up.

Yes, in the testing results section, thank you. So these are scores on a five-point scale, and generally they scored four plus on usefulness, interpretability, accessibility, recommendations for use. Questions, comments,

discussion?

MR. WILLIAMSON: All right. We'll start the voting here for validity. You have a minute.

And we've got them all. For validity, we have 12 high, 6 moderate, 0 low and 1 insufficient evidence. All right, so we passed the scientific acceptability.

CO-CHAIR BRISS: In terms of usability, the subgroup generally thought that this, maybe not surprising, thought that this had high usability both for quality improvement and for reporting.

CO-CHAIR PINCUS: Any comments, questions? Oh, Vanita?

DR. PINDOLIA: So Helen, is this where we would consider how we can move it into a composite versus having it single, individual or no, that's not something we can --

DR. BURSTIN: No, that would be later. Basically, our process is to evaluate each individual measure, and then we'll come back to

any of those discussions afterwards.

CO-CHAIR PINCUS: One question I had, just a clarification of the statement under 3(a)(2). When the respondents were rating this, were they rating this individually or were they rating this overall across the use as a suite of four measures?

MS. LAWLER: They were rating it as an individual measure.

MR. WILLIAMSON: Okay. We will now begin voting for usability.

We're still waiting on three responses. Everyone please make sure you point it at -- all right. We're now at 19. For usability, we have 15 high, 2 moderate, 2 low and 0 insufficient.

CO-CHAIR BRISS: And briefly, in terms of feasibility, the work group generally scored this as high or moderate. So the good news is that this is pretty routinely collected in clinical care.

The bad news is that the measure is not

yet specified and all of the component parts, as I understand it, are not easily electronically accessible. So work group assessment was generally high or moderate.

CO-CHAIR PINCUS: Any comments, questions on feasibility?

(No response.)

CO-CHAIR PINCUS: Okay. So let's vote.

MR. WILLIAMSON: Okay, we will now begin voting on feasibility. Begin now. And the results: we have 10 high, 8 moderate and 1 low.

CO-CHAIR PINCUS: So now we come to the overall vote for suitability for endorsement. Did you want to make any sort of overall comments on that?

CO-CHAIR BRISS: I don't think so. I don't have anything else to add that we haven't talked about already.

MR. WILLIAMSON: Great. We will now be voting on the overall suitability for endorsement. Begin voting now.

And for overall suitability for endorsement, we have 16 yes and 3 no. The measure passes.

CO-CHAIR BRISS: And congratulations. We exceeded Helen's 90 minute expectation.

DR. BURSTIN: We're half an hour over, but we're a half an hour under what it usually takes committees to do their first measure. So you're on track.

CO-CHAIR BRISS: We were faster than anticipated.

Measure 1654

MS. FRANKLIN: So that brings us to Measure No. 1654, TOB-2, Tobacco Use Treatment Provided or Offered, and a subset measure, TOB-2A, Tobacco Use Treatment, and we'll have a statement from the developer outlining this, and then we'll go to the lead discussant.

MS. LAWLER: Well, this measure again is based from the first measure for Tobacco Use Screening, where we find out whether or not the

patient is using some type of tobacco product. So when we find that there is a positive screen and they are using, there is to be a -- not a brief intervention, but counseling for the patient.

So there's a bedside counseling between the patient and the health care provider, and there are certain components of that counseling that need to be done. So we look to be sure that all facets of the counseling are completed, and then also part of this measure is that not only will they receive the counseling, but they should receive one of the FDA-approved medications for tobacco cessation.

Again, the denominator is just the patients that were screened positive for using tobacco products; the numerator is those patients that receive the counseling and one of the FDA-approved medications.

Oh. One thing to note here in this measure is that people that refuse the counseling or one of the FDA-approved medications will flow to the numerator. That's

because we give the hospital the benefit of doing the right thing. They tried; the patient didn't want to have the counseling or didn't want one of the medications. We're still going to flow that case to the numerator.

When we did an analysis of our pilot data, because there are so many combinations that can go on, and you get the counseling, but you don't get a medication, or you get both or various combinations, we wanted to know what was sitting in the numerator.

And we felt that for reporting purposes, it was going to be important to have some transparency, that people really needed to know those people who actually received the treatment.

So that is the reason for the subset measure 2(a). So both would be reported, but the subset measure is just those who actually received the treatment.

Also, let me just make note too that there are certain populations of patients that

don't need to have the medications, and those are the pregnant women, those that are using smokeless tobacco products, and there's one with light smokers.

That's why we ask volume in that screening measure, so we can find out if the patient is smoking, so that if they are a light smoker, we can exclude them from receiving the medication. So those folks will need to get the counseling, but they don't need to get the medication.

MS. FRANKLIN: Any other comments from the developer table?

DR. FIORI: Just very briefly to add, that the evidence base for these two components of intervention, counseling and FDA-approved medications, is very substantial for each of them individually, and with independent analyses of the data, including both Cochrane in the 2008 Public Health Service Clinical Practice guideline.

The evidence also supports the use of

both components, counseling and medication. Each of them are effective independently, but there is an additive effect of combining counseling with medicine, and that's why it is listed as such.

MS. FRANKLIN: Thanks. Actually, we will have Caroline lead the discussion, and then we'll have questions about this. Or did you have something related? Oh, okay, okay. Go ahead.

DR. CARNEY-DOEBBELING: That wasn't me.

(Laughter.)

DR. CARNEY-DOEBBELING: But it sounded like it came out of my mouth, the first sound effect of the day. The measure, as described, is to look at those folks who were found to be smokers, and from that group, which of those individuals received counseling and medication, or refused counseling or medication, and subset of that, of those smokers who received counseling and medication.

There are some interesting findings as we worked through this, and should we follow that?

MS. FRANKLIN: Yes. We should start with the importance discussion.

DR. CARNEY-DOEBBELING: The importance, as stated earlier, tobacco use and the negative outcomes of tobacco are well-known, and the importance of screening and offering treatment for tobacco cessation is not in argument based on the current evidence.

CO-CHAIR PINCUS: Any comments or questions with regard to the issue of impact?

(No response.)

CO-CHAIR PINCUS: Okay. So we're prepared to vote?

MR. WILLIAMSON: We will now be voting on impact. Begin voting now. That was quick.

DR. CARNEY-DOEBBELING: No arguments on impact.

MR. WILLIAMSON: We have 19 high, 0 moderate, 0 low and 0 insufficient.

DR. PATING: Could I ask just a clarifying question? With regards to the intensity of the counseling, is there -- I was trying to look --

DR. CARNEY-DOEBBELING: We'll get to that. That's part of what I'll discuss in the next section.

CO-CHAIR PINCUS: Okay, performance gap.

DR. CARNEY-DOEBBELING: When this measure was built, it was built on top of the prior measure. It's the second measure in the suite of measures being offered by the Joint Commission.

At this point, they have looked at of those individuals who were screened positive and provided data in hospitals, that the rate of individuals who are identified and then go on to be offered treatment is quite low, ranging in some cases from as low as 16 to as high as 35 percent.

Even in populations of folks who have

had an adverse outcome from a tobacco-related illness, the rates of being offered treatment are quite low. The gap, according to the literature, is significant.

CO-CHAIR PINCUS: Any questions about the performance gap?

(No response.)

MR. WILLIAMSON: Okay. We will now be voting on the performance gap. Begin voting now.

Okay. We have 18 high, 1 moderate, 0 low and 0 insufficient.

CO-CHAIR PINCUS: Let's move on to evidence.

DR. CARNEY-DOEBBELING: The evidence for the intervention being measured with this particular measure is not quite as strong as originally stated. According to the literature, there is very little work done specifically looking at hospitalized inpatients.

The 2007 Cochrane analysis focused

primarily on inpatients who had received contact plus at least one month of follow-up associated, and had a more intensive counseling intervention than what is offered, particularly with this measure.

The bulk of the other analyses is in the outpatient setting, and is not applicable to the measure under discussion. There was no further evidence in the analysis that linking the medical condition of the member along with the tobacco counseling resulted in any higher likelihood of quitting.

So from that point of view, I think that the evidence supporting the measure in the inpatient setting is low. Further, the panel was asked to provide to this group the quality of the body of evidence with regard to RCTs. Those were to be submitted on a table, and the table was lost and never produced to this group.

However, the panel said that the table had shown a level A grade. However, the direct evidence for RCTs, according to what was

submitted, was not provided.

CO-CHAIR PINCUS: Jeff Susman.

DR. SUSMAN: So are you arguing for insufficient evidence? Is that essentially what you're saying?

DR. CARNEY-DOEBBELING: I will get there, but that's where this is heading.

DR. SUSMAN: Okay, thank you.

DR. CARNEY-DOEBBELING: The rest of the evidence provided is based on United States Preventive Services Task Force guideline endorsement, that counseling and treatment should be offered, but no further findings or publications other than that were supported.

And so my overall rating for the body of evidence being used to generate this in the setting of inpatient hospitalization is low.

That is complicated by, because if we don't get to the next step about the intervention itself, and I'd like to address what I think was David's earlier question.

The intervention itself calls for a

clinician. The type of clinician is not defined in the measure, whether that is a physician or other health care provider, to participate in the counseling and treatment suggestion that's not specified in the measure.

This also provides some issue with the evidence at hand, because that has mostly been studied when physicians and physicians only are providing the intervention and the counseling.

The second is that the type of counseling that is required by the measure is tobacco use treatment practical counseling, and the practical counseling must involve three separate components, and those components must be documented in the medical record, that there is a recognition of danger situations, that there is assessment of developing coping skills, and thirdly, to provide basic information about quitting to the member and referral to say, a quit line, is not considered adequate.

The medical evidence in the past looks at the amount of time spent in counseling, and

there is a high correlation between spending one to three minutes, I think it's three to ten or more than ten minutes, each of those having a much higher association then with downstream quitting ultimately.

So this, I think, becomes an issue because the evidence suggests that it needs to be physician-led. There is a lot of documentation that would be included in this, and whether or not all components are done routinely and documented routinely is something that we need to take under consideration.

CO-CHAIR PINCUS: Caroline, could you maybe go a little bit more into your thinking about low versus insufficient, in terms of how you thought about it?

DR. CARNEY-DOEBBELING: At present, the evidence would be insufficient, because there simply has not been enough work done in the inpatient setting. Further, there has not been enough specific work done on the type of counseling that is being required by this

measure. So I would suggest insufficient as opposed to low.

CO-CHAIR PINCUS: Could the developer comment?

DR. FIORI: I'm happy to respond, and what I'll be using as my response basis is the 2008 United States Public Health Service Clinical Practice guideline, and some review of almost 9,000 published manuscripts.

In there, there is about 100 meta-analyses that for inclusion in the meta-analyses requires it be an RCT with a bunch of other specific criteria. So to the issue of data supporting counseling beyond physician, in fact I would respectfully disagree and say that there is substantial data for counseling provided by non-physician clinicians.

In fact, the meta-analyses and the PHS guideline included all clinicians in those meta-analyses and did not restrict them to physicians. We did not find a difference when we did subanalyses of quit rates for counseling,

based on whether it was a physician provider or a non-physician counselor.

So at least in the PHS guideline work, we found that physician and non-physician counseling was effective in boosting cessation rates. To the really core question of is there substantial enough data in hospitalized patients to provide a recommendation for counseling, I guess I'd answer it in two ways.

First, there are some studies, and surely not to the degree there is in the outpatient study, but there are some studies, and those have consistently shown an effect of counseling in driving downstream quit rates. But I would also suggest that counseling provided irrespective of the setting, if provided in an evidentiary-based way, should have the same effectiveness.

We have not seen an analyses for smoking cessation counseling at least, that the setting changes the downstream quit rates in general. It is positive irrespective of the

settings, and we've looked in settings based on dental offices, clinician offices, group counseling settings, individual phone counseling.

In essentially every place we've looked, we've gotten a same consistent finding, that counseling for smoking cessation boosts success rates. So maybe I'll stop there.

DR. CARNEY-DOEBBELING: I think the argument that I was presenting is not about whether any one of those individually would be successful, but that they are linked, that the type of clinician and the measure wasn't specified, that it was all clinicians, physician, non-physician, whatever it might be. So some room for clarity in the measure with that.

Secondly, because we're talking only about the inpatient setting, what was provided in your submission didn't really focus on the inpatient setting, other than to say that the Cochrane review had little evidence to -- had

some evidence to support that, but that there was not much in what you provided on the inpatient setting only.

Thirdly, that the type of intervention which is specified has in and of itself met specific, was not included in what you submitted any data, that that specific type of counseling was successful, what the downstream consequences of that were, and whether or not it can be routinely conducted well in the inpatient setting, and evidence later in this will talk about that under reliability.

CO-CHAIR PINCUS: Questions, comments from the panel. So I have Jeff Samet.

DR. SAMET: Just a clarification request. The U.S. Preventative Task Force in this realm, it wasn't clear. Was that all settings or was that hospital-specific, the recommendation?

DR. CARNEY-DOEBBELING: It's all settings.

DR. SAMET: Okay.

CO-CHAIR PINCUS: Peter and Dodi.

DR. KELLEHER: I'm looking at the overall work group individual evaluations of this measure, and I was wondering if we could go through them, because I'm interested. Not everyone agrees with the lead discussant, so I'm interested in why people varied from her position.

DR. CARNEY-DOEBBELING: In the overall work group for scientific acceptability, only four members participated in the overall work group. There were two highs and two moderates. I'm sorry, that was scientific acceptability on the quantity and quality. Those were each 3 highs, 1 low and 1 moderate.

CO-CHAIR PINCUS: Peter and then Lisa.

CO-CHAIR BRISS: So I guess I have a question. It's probably true that the data that are presented are -- leave something to be desired in terms of breaking out subset analyses

of the whole body of evidence by setting and type of provider.

But we're talking about a couple of generally high quality reviews with more than 30 trials, right? So for the evidentiary standards in this system, you know, to get to moderate evidence, you would need three consistent trials, as a general rule, right?

So I wonder about it's a little hard to believe that among this huge body of evidence that there might not be at least moderate evidence on this topic.

DR. CARNEY-DOEBBELING: And I'm not disagreeing with that, only I do need to give a caveat. I'm sorry. I was looking at the wrong page. I hadn't flipped my page to 1654 when I said that there were only four work group members. I did not have the opportunity to participate in the work group because of some glitches in notification.

So my comments are made outside of whatever discussion happened at that work group,

and I am giving a rating based only on what the measure states, which is linked only to the inpatient setting. So if we want to broaden above and beyond what the evidence is for this measure, that's the will of the team.

The measure itself is limited only to the inpatient setting, which is what is driving my comments. I have no quibble whatsoever that counseling is effective, and I want to go on record saying that only what the evidence is that directly supports this measure.

CO-CHAIR PINCUS: So I have Lisa and Jeffrey and then back, Dodi, are you, do you have --

DR. KELLEHER: Well, I just, so that was -- could we go over what the updated evidence is from the work group, I mean evaluation is from the work group at some point?

(Off record comments.)

DR. CARNEY-DOEBBELING: Well, if any of the work group members would like to comment.

CO-CHAIR BRISS: So the rest of the

work group -- on this topic, the rest of the work group was --

DR. CARNEY-DOEBBELING: High or moderate.

CO-CHAIR BRISS: Yes, was friendlier to this body of evidence than Caroline was. So in terms of quantity, 5 high and 1 moderate; in terms of quality, 4 high and 1 moderate. 1 low, I'm sorry.

DR. CARNEY-DOEBBELING: Again, I want to say that I am not quibbling whatsoever with the fact that counseling works, only what was provided with regard to evidence in the patient setting.

CO-CHAIR BRISS: And consistency was 5 high and 1 low. So I guess there could also be, you know, what counts as direct evidence is also, it seems to me to be a little complicated here.

So at some point, at some point, I think that I would read this evidence that eventually once you proved counseling works in

some number of settings, I'm not sure that I would require it to be reproven in every conceivable setting.

CO-CHAIR PINCUS: Lisa, and then Jeffrey Susman and then Tami Mark.

DR. SHEA: I just had a question or clarification regarding counseling. Is that individual, or could it be done in a group and is specific to individual counseling?

DR. CARNEY-DOEBBELING: The way that the measure currently reads doesn't specify either/or, but suggests that it's individual at the bedside counseling.

DR. SUSMAN: It seems to me we're going to deal with the issue of transferability of interventions or screening from one setting to another amongst many of these measures. It also seems to me in this case, with tobacco use counseling and drug intervention, we have probably as robust a body of evidence as almost anything we're likely to see.

For me, the overall, you know, sense

is that it's a reasonable leap of faith to say that all this Cochrane studies that were reviewed, the U.S. Preventive Services Task Force, etcetera, suggests that yeah, you know, it's not perfect, but seldom do we have perfect evidence. For me at least, this is more beyond good enough.

DR. MARK: My understanding is that they have to receive not just the counseling, but also the pharmacotherapy, and so I had a question about that.

Looking at the write-up, it says that based on the USPSTF, the rate the risks of pharmacotherapy as small, but these medications do have a black box warning from the FDA, for serious psychiatric potential side effects. So my concern would be, you know --

DR. CARNEY-DOEBBELING: Some of them do, some of them do, and there's a wider, a broader range of products t hat are acceptable under this measure. Back to the counseling, according to the measure, that counseling has to

be --

CO-CHAIR PINCUS: Wait, wait. Caroline, I just -- I'm not sure Tami finished her comments.

DR. CARNEY-DOEBBELING: I'm sorry.

DR. MARK: Yes. So my concern is that, you know, a physician who would think that the person's, there would be contraindications due to the psychiatric state or perhaps due to some other medications that were taken that would cause interactions, you know, would discount those in an effort to, you know, comply with this quality measure. So I'm concerned --

DR. CARNEY-DOEBBELING: And people can be discounted if it's documented, why they were discounted, or if there was a reason for not going forward with the medication component of the measure. Nicotine replacement also is an allowable medication in that component of the measure.

So some of those drug-drug interactions or potential black box warnings may

not be included. Women who are pregnant, people who are cognitively impaired, light smokers and those who use smokeless tobacco are removed from this measure as well.

DR. MARK: Yeah. I heard that the denominator was going to remove women who are pregnant, adolescents and light smokers, but I didn't hear that removed from the denominator are any patient that the physician thought, you know, had contraindications to pharmacotherapy.

DR. CARNEY-DOEBBELING: They're not automatically removed from the denominator, but they can be if there's -- an exception can be made to treatment if there is a reason why treatment should not be given.

CO-CHAIR PINCUS: Is that a formal part of the measure?

DR. CARNEY-DOEBBELING: I don't read it as a formal part, but somewhere in the text that came out.

CO-CHAIR PINCUS: Let me just ask the measure developer. Is acknowledgment of

contraindications a formal part of the measure specifications?

MS. WATT: Yes, that is true. So patients -- the data element is patients with reasons for not administered FDA-approved cessation medication.

DR. FIORI: And five of the seven medications are nicotine medicines that are approved by the FDA. So five out of the seven are nicotine.

CO-CHAIR PINCUS: Tami, do you have anything else? Bonnie?

DR. ZIMA: I had really a question, I think, for Helen. When Dr. Susman raised his point, does that mean that he's asking for an exception?

DR. BURSTIN: It's a good question. I think this is one of those gray areas of how much it's okay to extrapolate from a very, you know, substantive body of evidence, I think it's fair to say, around tobacco cessation interventions and outcomes, and I think I think

Dr. Fiori explained this well.

There's certainly evidence in lots of different settings. I'm not sure I've heard anybody invoke why it would necessarily be different on the inpatient side versus the outpatient side. So I'm not sure I see it as indicating where there's --

The exception is really where there is not evidence, and therefore you're invoking that you think the risks would be significantly lower than the benefits. I think this is very much a gray area, since the studies are so mixed in all settings.

It's not as if there's not evidence for the hospital setting. It's just that it tends to be the evidence is for all settings.

CO-CHAIR PINCUS: Let me ask for a point of clarification, and I think just going forward, it would be very helpful if measure developers, when they fill out these forms, if they're talking about a specific setting, they actually cite evidence, even though there are

generalizable things, so that we can at least see what's the subset of evidence specific to the hospital setting, so that at least we can put that into context.

So just among -- do you know, among the various studies that were cited, what number were in the inpatient setting?

DR. FIORI: Well, we've got evidence for counseling and a separate body of evidence for medication, and the number of studies that went into this actually is extraordinarily robust --

CO-CHAIR PINCUS: No, no. We just want to know what is the number of studies?

DR. FIORI: So I was going to say that there's more than a handful at least on both, and I would think that it might be on the medicine side, as many as 20 or more studies that have been tested in medical, in inpatient settings.

On the counseling side, I think it's fewer, but I can't tell you the exact number. I could attempt to get that, but it will take me

a little bit of time.

CO-CHAIR PINCUS: Well, I think it would be useful to sort of track the information on a consistent basis, so that we can evaluate it, because that's clearly coming up as a key issue across these various measures that we're dealing with. -- specific to the particular setting in which it's being applied.

I can't imagine that there may be differences -- very directly, in terms of the degree to which patients are actually paying attention to it and focusing on it, because of other obvious distractions that are going on during a fairly, what has become largely very short inpatient stays, as well as, you know, on the medication side, whether people, even if it's prescribed in the hospital, whether it continues afterwards.

All of those, you know, we know how often -- between what happens in a hospital and what happens later.

DR. CARNEY-DOEBBELING: I also want

to reiterate that the measure not only specifies the inpatient setting, but prescribes specifically what type of counseling must be done, and that the three elements of that type of counseling must be included in the documentation.

So while the broader body of evidence suggests that counseling is effective, this measure specifically narrows it down to tobacco use treatment, practical counseling, of which the recognizing danger situations, developing coping skills and providing basic information about quitting must be documented.

DR. FIORI: And the basis for that, just to share, is that the Public Health Service guideline panel analyzed components of counseling, and identified those as the ones that were components of counseling that specifically resulted in downstream quit rates.

CO-CHAIR BRISS: Yes. So I thought that the Cochrane -- so I asked some clarifying questions. So one of them is how many studies.

The Cochrane review, as I understood it, was actually limited to health care setting, I'm sorry, to inpatient settings and to hospitals.

So the number that's associated with that review in the submission is 33. So if it's -- so it sounds like there were at least 33 studies in hospitalized patients, which I would generally consider to be a huge body of evidence, right. I'm sorry --

DR. CARNEY-DOEBBELING: But again, it was -- and I'm reading directly. "The 2007 Cochrane analysis revealed that intensive intervention, which was inpatient contact plus follow-up for at least one month, was associated with a significantly higher quit rate." That's not the same as what the measure is calling for.

So I might be being too literal in my application of the evidence, but I'm trying to link what was provided to what the measure is actually asking for.

CO-CHAIR BRISS: Yes. So I was on the work group and was among the people that was

friendlier to this measure, and so my rationale for this is that I myself would not require me proving in every conceivable setting, right?

I also think that there's a -- so the thing that we're sort of hung on, it seems to me to be a generalizability argument, at least in part, and it's what body of evidence might bear on this measure. So as I hear it, there are folks who would say I would like a specific body of evidence, sort of limited specifically to this counseling, as defined in the measure, and delivered in the hospital setting.

I would have said that given the body of evidence that we're drawing from, that I might not require either of those in quite so specific. I would say that there's a fair body of evidence across settings that suggests counseling works, and there's a fair body of evidence across settings that suggests that more intensive and longer -- that any counseling works, and that more intensive and longer counseling works better.

So I would have said that all of that body of evidence could probably be brought to bear on this measure. I understand that it's possible for reasonable people to disagree about that.

CO-CHAIR PINCUS: Are there any further comments before we vote? Jeffrey.

DR. SAMET: This may be helpful. It's intended to be. You know, as far as the setting, when you think about settings, because we'll see this again, conceptually is the setting different from the outpatient setting? We're talking about inpatient here.

The other piece is so is there something about the setting that's unique, that makes it normal or less? The other piece is are the people identified in that setting different, inherently different in the inpatient setting than the outpatient setting?

So those are really two different issues that play into this, not being generalizable. For smoking, my perspective is

that the latter, they're not, because the people who smoke are going to be at a dependence level in either case. It's not like it's milder than that.

So I'm more reassured, kind of from that theoretical basis, in addition to the argument that there's that type of data. But that's all. That's my perspective.

DR. BURSTIN: I just wanted to give clarification from the Cochrane review. Just I think this was the point Caroline was trying to get at, in fact that it was actually only the combination of the inpatient intervention plus 30 days that was affected.

Inpatient alone has not been shown to be effective. They say interventions of lower intensity or shorter duration. So I think that's what Caroline's getting at here. So to me, it seems more of an issue of it's really the combination of Tobacco 2 and Tobacco 3 together --

CO-CHAIR BRISS: Or Tobacco 2 and

Tobacco 4 --

DR. BURSTIN: Right, that probably is really the sweet spot perhaps.

CO-CHAIR PINCUS: Any further comments? Jeffrey. Let's go in order.

MS. HOO: Just to follow up on Dodi's original question, and echoing Peter's comments, in the work group call, I think there was a general sense that some of the research was generalizable to the inpatient, and I think where the work group had concerns was more on the feasibility and usability, that given in an inpatient setting, there's probably a lot of other issues going on, and similar for the next measure, that at discharge there's a whole other set of lists of issues, medication management, all sorts of things, that this may not be at the top of the list.

But I think there wasn't so much a quibble about the evidence and the ability to generalize the effectiveness of counseling, and that this indeed would be a teachable moment for

a lot of inpatient individuals.

CO-CHAIR PINCUS: Nancy.

DR. HANRAHAN: Yes. I just wanted to say that's exactly what I was thinking too, that where there's this gray area here between the evidence that we're considering and how it crosses over various settings, and the usability and feasibility.

So I just got that clear from that, from what you just said, and I think that's important.

CO-CHAIR PINCUS: Tami.

DR. MARK: Following what Dr. Samet said, I would be concerned about again, the population in particular, the fact that they're complex medically and now you're adding another medication on top of people who are going to be taking lots of medications and having lots of issues. I think that's dangerous. Your thoughts on that. Or interested in whether the evidence speaks to that.

DR. SAMET: You know, I don't think I

can answer that with an evidence-based response, which is what we want to use here. I think it's a legitimate concern. I know I can tell you in a clinical sense we don't really care about that.

(Laughter.)

CO-CHAIR PINCUS: Does a developer have a response to that question?

CO-CHAIR BRISS: Right. The question, essentially as I heard it, was how do you think the evidence generalizes to complicated or distracted patients?

DR. SAMET: In terms of adverse events, she asked.

DR. FIORI: Maybe two things. The data, as I've reviewed it, applies across patient settings, including complex patients.

Some of the earliest smoking cessation studies in fact were done back in the 80's with people in CCUs, right after an acute MI, and found an extraordinarily high rate of cessation if you provide counseling at that teachable moment, right after an acute MI.

CO-CHAIR PINCUS: And this question had to do with medications.

DR. FIORI: No, and to the issue of medications, I think it's very well-taken. Of course, using nicotine medicine is something the patient is already tolerating and ingesting.

So if the physician was concerned about drug interactions, and also because the nicotine medicines have immediate onset, whereas these other medicines take up to a week to have effectiveness, in inpatient settings it's almost exclusively nicotine as the medicine used, and we already know the patient tolerates it.

We're just giving nicotine without the other 4,000 chemicals --

CO-CHAIR PINCUS: But just a clarification then. Then why does the measure include non-nicotine mitigations?

DR. FIORI: Because I think there is -- there could be a place for them, and all seven of these have been shown to be effective.

Another reason why one might want to use one of the pills is if the patient has already used on the nicotine products, and has reported that they're ineffective.

There also is some data that one of the pills, varenicline, is more effective than the nicotine products by themselves. So there are a couple of reasons why you might, in practice, the bulk of the use is nicotine medicines in the in-hospital stay.

DR. SUSMAN: This is a process issue, I think, for Helen. You alluded that we're going to obviously consider each measure individually on its own merits.

But after we're all done, is there a point to weigh in and say perhaps one of the intervention measures is better than another, and then looking at what other existing measures might look like?

DR. BURSTIN: We'll certainly have an opportunity to look at anything that you think requires harmonization, or measures that might

be competing. It's unusual, but the question is could the same developer have measures that are in fact competing, and I think that's a question I think we'll need to better understand the approach of are these in fact paired IE, and in our parlance, a paired measure is actually a part of endorsement.

Meaning you should only ever see Rate A with Rate B, or are these really just the suite, which is more general. I think those are questions we'll have to grapple with after we go through each of the individual reviews.

CO-CHAIR PINCUS: So I think we're ready to vote. Is that okay? Okay. So let's vote on the sufficiency of evidence.

MR. WILLIAMSON: Okay. We will be voting on evidence. Begin voting now.

CO-CHAIR PINCUS: Right. Yes, no or insufficient. This is a 1, 2 or 3. There's no 4.

MR. WILLIAMSON: If you need to revote, if your first vote was no, you can just

press the number for your final vote, and that's the one that we'll record.

Our final vote, we have 16 for yes, 0 for no and 3 for insufficient evidence.

CO-CHAIR PINCUS: So let's talk about reliability.

DR. CARNEY-DOEBBELING: The next section under review was reliability. The work group voted 4 high, 2 low. I'm sorry, 2 moderate and none low. According to the data submitted by the Joint Commission, the study of reliability showed a kappa of 0.113.

There were 25 percent false positives and a significant 38 percent, I think it was, disagreement between inter-rater reliability in the original sample and the review sample.

One reviewer in the initial group had brought up that agreement levels for counseling are not ideal, and the false positives, the low inter-rater reliability score and the kappa were driven by the issues around counseling and documentation of counseling.

CO-CHAIR PINCUS: Comments, questions about reliability?

DR. CARNEY-DOEBBELING: I can give you some more specific numbers. 131 cases were re-abstracted. 25 percent or 33 of those resulted in a false positive.

The primary reason for false positives was related to the agreement of only 38 percent for the data element tobacco use of the counseling, and it was because hospitals often gave credit for counseling, when not all of the components of counseling were given.

The developers went on to state that they felt that over time, people would learn what this was and would go on to do it correctly. But that has not been tested.

CO-CHAIR PINCUS: Any comments, questions? Some people have their cards up. Peter.

CO-CHAIR BRISS: So in addition to the -- so the numerator's clearly hard here, right. In addition to the numerator, another thing that

happened on the work group call was there were denominator questions. So I was actually a little concerned about how the denominator got defined.

So essentially it's -- what gets into the denominator is people who are tobacco users and you're excluded from the denominator if you're not a user, or if the system doesn't know and didn't assess. So we know from the previous measure that as much as 40 percent of people didn't get assessed, based on the literature review.

So the question to the developer is why do we let people off the hook on this measure for not assessing. So --

MS. LAWLER: Well first of all, we felt like it would be difficult to put into the population patients where we didn't know their status, whether or not they were tobacco users. Again, I would tell all of you this is a proportion measure. Can you hear me? I see you guys going like this. Can you all hear me?

Okay, okay, better.

So being a proportion measure, where the numerator is a subset of the denominator, and every case has an opportunity, equal opportunity to move to the numerator, we don't want cases in there that are not going to have that chance to move to the numerator.

So when we don't know the tobacco use status, we've excluded them. I think there was some concern about perhaps gaming, so let's just not screen the patient and then -- and we've coupled that value with a UTD, unable to be determined.

So there was concern about that, I think, in the work group as well. We did do, get from our statistician some data. So for tobacco use counseling, those that where the counseling was not offered. Let's see. MS. WATT: There were a total of 2,598 cases that we were looking at. It was not offered in 286, and unable to determine in 155.

DR. CARNEY-DOEBBELING: I had a

question for the developers about the ICD-9 coding that's part of the algorithm. Is that for the primary reason for hospitalization, or that there was an ICD-9 code for tobacco use or dependence. I was confused with the algorithm.

MS. WATT: I didn't hear the first part of the question. Could you restate it please?

DR. CARNEY-DOEBBELING: Sure. In the algorithms for identifying cases for review during the accreditation period, part of that is based on ICD-9 principle diagnosis codes. Are those for checking for the reason for the inpatient stay, or checking that tobacco use or abuse or dependence was coded?

MS. WATT: No. That was checking for pregnancy, so we can exclude those patients that are pregnant from receiving the medications. That's what the ICD-9 code is for there.

DR. CARNEY-DOEBBELING: Thank you.

CO-CHAIR PINCUS: Okay. Actually, I want to call on myself, to step out of the chair

role. So a question. The kappa of .113 is considered poor by most standards, and just -- and I understand part of the reasons for that. But still, it still results in an overall poor reliability, and at the same time, I'm just wondering about how, particularly on the medication side, there's a lot of specific components that are required to be assessed, including whether they're light smokers, whether there's contraindications and all that.

And I'm wondering just do we know anything about the reliability of those particular items, in terms of their assessment?

DR. CARNEY-DOEBBELING: They reported 68 percent for the data element reasons for no tobacco cessation medication, and 73 percent for the data element tobacco use treatment FDA-approved medication. It was the 38 percent for the counseling that drove that kappa.

CO-CHAIR PINCUS: Peter.

CO-CHAIR BRISS: So this one's

different from the last one, it seems to me. So the last one had a poor kappa that I could explain, and this one has a poor kappa that in part reflects a lot of disagreement about what goes into the numerator, right.

So the hospitals were, it sounds like the hospitals were less demanding about what counts as counseling essentially, than the Joint Commission was.

So my question is sort of a general one about this strikes me as being likely a problem that we could have in any clinical interaction, where you're trying to capture what the clinician did in a setting that doesn't generate some hard data like a laboratory test, right.

So I don't know. I don't know if Helen or somebody can give us the sense of how reliable generally is essentially a clinician self-report of in the room, I did the right thing, and how does this compare with the rest of the universe of similar measures.

In a world where I watch my residents

on every Friday sort of cut and paste stuff into EPIC, that I know doesn't really -- and honest, this is real problem, right? They cut and paste stuff into EPIC that I know they didn't do. So how does this compare with the rest of the universe of similar measures?

DR. BURSTIN: It's an excellent question. I'm not sure I can answer it, other than saying I think it's a little different if you have clinician self-report. Obviously, you're going to have much higher reliability if you're self-reporting on what happened.

I think the issue here, I'd be curious to know, you know, perhaps from the Joint Commission, who has been doing hospital paper-based assessments like this for many years, how does, you know, a 38 percent agreement rate for the main data element compare, perhaps to some of the other assessments of things that don't relate to drugs, or some of the things that are much harder that you can always find in data? I mean that's, I guess, a question for you, Ann.

You probably know that best.

MS. WATT: I'll speak first, and then if Nancy has anything to add. You know, 38 percent we're not thrilled with obviously, which is why we, you know, strengthened the specification to say you have to have all three of these things.

Frankly, when we were on these reliability visits, people said yes, we told them, and they sort of had overlooked, apparently, the specification that says oh, and it has to have these three components. Maybe we didn't word it as elegantly as we could have in the submission, but one thing we have learned is that hospitals learn the specifications over time.

Frankly, that's what we chalked this one up to. They abstracted what they wanted to believe, not necessarily what they were asked to believe based on the specifications.

MS. LAWLER: I think that's correct, and I think too, you know, you always have a

tendency. You're abstracting, and maybe you were the one who did the counseling, and yes, I know I did that, but it's not there and it's not documented.

The other thing is that I think a lot of people were used to using our old smoking cessation advice and counseling measures, which were not nearly as fulsome as these. You could just hand a brochure to someone and they didn't even have to look at it, but you got credit for doing the counseling.

So for years, people had been doing that, and in fact that measure was still being used as the time we were introducing these new measures.

So I think people were just really used to doing not quite as good a job at the counseling, but so were giving themselves credit for it, and we were requiring a lot more.

CO-CHAIR PINCUS: Any comments or questions with regard to reliability? Nancy.

DR. HANRAHAN: Yes. I think the

reliability again is related to the usability and feasibility of this particular measure being used in a hospital setting, and I'm speaking for all nurses that work in hospitals. As you have, Dr. Briss, spoken for physicians, these kinds of measures are very difficult to really implement in the real world.

Yes, clinicians do, you know, dub in or copy in. I don't think that is a reflection, or I would say it's not a reflection of their intention to help people stop smoking. So I just, I have real trouble with this measure from the reality, the real world perspective, and whether or not it's going to be utilized in the intended way.

DR. KHATRI: That there are multiple factors that play a role in whether or not an intervention is implemented, not just, you know, the EHR and time and other competing demands, but also patient readiness to change. So there's an interaction there.

So I can imagine that this

intervention would be much more easier if a patient shows that they're ready, engaged, want to be part of it, versus "I don't want to hear it, don't talk to me about my smoking, I don't care about that now."

And then so that point, that also limits the clinician. So I think we have to be aware of -- it's kind of very multi-factorial, in terms of how this can be measured and evaluated. It's not just the clinician wanting to do the right thing.

CO-CHAIR BRISS: I'm afraid on the data on this one, this one's starting to feel like low to me. On the feasibility issue, the only point that I would make that if we -- I'd hate to have us eventually wind up sort of writing off all the medical things that sort of require human interaction, because I think it gets us to a real, in a real bad sort of drunk at the lamp post problem, where we're only looking at things that are easy to measure.

And I don't think that -- I think that

this may be one of those really important things to do, that we shouldn't write off just because it's kind of hard to measure.

MS. WATT: As the developer, could I make just one comment? That people, to keep in mind that people that do refuse the counseling flow to the numerator of this measure again. So they do have an opportunity, and as a clinician, you're not going to --

When they tell you I'm not ready. I don't want to listen to this right now, then, you know, that's okay. There's still, again the case will flow to the numerator. In the subset measure, of course we'll only see those that got the actual counseling. So I just wanted to bring that point forward.

DR. SUSMAN: So I think we have another rating on usability, that covers a lot of the issues that we're merging into now, and I think we should probably stick to the question at hand, just an observation about the discussion. I don't disagree with many of the

points that have been raised.

CO-CHAIR PINCUS: Ready to vote on reliability?

MR. WILLIAMSON: Okay, we'll be voting on reliability. There we go. You may begin voting now. Still waiting on one vote.

All right. Now we're good. Okay. So we have 0 high, 8 moderate, 7 low and 4 insufficient evidence. So this measure does not pass scientific acceptability.

CO-CHAIR PINCUS: Are there any comments on --

DR. BURSTIN: Well, we have to pass validity. So just to be clear, what we're talking about is on your quick sheet you've got there, page three, the decision logic is you must have reliability and validity rated moderate or high to meet this criterion. So technically, it's done.

CO-CHAIR BRISS: And I would have said you can't you can't have a valid measure if it's not a reliable measure. So I wonder if we need

to actually have the discussion before we turn this one back.

DR. BURSTIN: The discussion of validity? Yes. I think it would be good to just finish up validity, so we could complete scientific acceptability.

DR. CARNEY-DOEBBELING: Validity was assessed for face validity, with three components, a public component, a survey component and a pilot site survey. Face validity of the original candidate measures was assessed through public content on a five point scale, relative to ten different characteristics. Slightly more than 2,000 persons were elicited.

A TAP survey looked at 11 members completed it. All the members were asked to participate in a TAP survey. Eight members completed t he TAP survey, and again that was a five-point scale on disagree, somewhat agree, neutral, somewhat agree and agree.

Both of those showed a validity score

for each of the different elements of measures, generally ranging between four and five. Those looked at the face facility of the measure itself. The TAP, for those who want more specification, was clarity, usefulness, interpretability, data accessibility and collection, and recommendation for national use or endorsement.

The group summary from that group didn't specifically comment about validity, unless I'm missing that.

CO-CHAIR PINCUS: So do we need to vote on it? Okay. So are there any comments or questions with regard to validity?

(No response.)

CO-CHAIR PINCUS: I just have a question, maybe to you Helen. But it seems to me that a lot of the validity statements are just about face validity, and I'm curious as to why people don't present data about the actual validity.

DR. BURSTIN: Face facility is one

form of validity. It's certainly not as high as we'd like to get. I think oftentimes, particularly in areas where there aren't other measures to know which the gold standard is related to, it's hard to know how else one would do validity testing in some of these newer areas of measurement.

CO-CHAIR PINCUS: Well for some of these, I can imagine number one is to look at sort of, you know, is it a cross-validity in terms of other similar measures, but also to look at, you know, if the measure is applied, are people less likely to smoke, because that would be the obvious way.

DR. BURSTIN: And again, that's very hard to answer.

CO-CHAIR PINCUS: Some of the data that you've described sort of underscore that, but to see if there's actually -- but one could imagine some prospective testing of that.

DR. BURSTIN: That's very difficult to do with a new measure. That's oftentimes

what we'd like to see at maintenance. Hard to do when a brand new is out there. We don't actually know what the experience will be going forward. We'd love to be able to have the developers gather that, but they often don't know that for a new measure.

CO-CHAIR PINCUS: Any other comments about validity?

(No response.)

CO-CHAIR PINCUS: So let's vote.

MR. WILLIAMSON: Okay. We will vote on validity. Begin voting now.

We're waiting on one more response. If everybody could please make sure you're pointed at the computer.

CO-CHAIR PINCUS: And we might want to move -- I wonder if the card you have is blocking.

MR. WILLIAMSON: Okay, all right. For the record, we have 1 high, 7 moderate, 7 low, and 4 insufficient evidence. So the measure fails on scientific acceptability.,

CO-CHAIR PINCUS: Okay. Should we

take -- we're running behind. Should we take a ten minute break? Okay.

(Whereupon, a short recess was taken.)

Measure 1656

CO-CHAIR PINCUS: So the measure we're about to do is 1656, and could we have a brief statement from the measure developer?

MS. LAWLER: We're on Tobacco 3, right?

MS. LAWLER: Okay. This is the third measure in our set, which again looks at only those -- in the denominator those patients that screened positive for tobacco use, and in the numerator, we're looking to see the number of patients that were referred to outpatient counseling, and were given a prescription for one of the FDA-approved medications.

Just like in the second measure, those patients that again refused the counseling and prescription will flow to the numerator. In the subset measure, you have just those patients that received the treatment, and we still have

the exclusions for those patients for medications, which were the pregnant smokers, the light smokers and smokeless tobacco users. So those three exclusions here from receiving the medication, but still needing to receive the referral for outpatient counseling.

MS. HOO: Sure. Do you want me to just go into -- in terms of the work group discussions, I think some of the issues that we discussed around Tobacco 2 parallel some of the considerations here. On the whole, at the work group call, folks felt that this had a high impact, and the voting was unanimous on that category.

CO-CHAIR PINCUS: Are there any comments or questions about impact? We've sort of been through this with each of the previous ones. Okay. Oh, Caroline. Oh, okay.

MR. WILLIAMSON: Okay. We're voting on the impact. Please begin voting now.

We're still waiting on one more response. Oh, we have one person gone. Okay.

We have 16 high and 2 moderate, 0 low, and 0 insufficient.

MS. HOO: With respect to the performance gaps, the voting in the work group was 4 high and 2 medium. I think that the in the studies that were cited, there were some differentials in terms of the stated results. Some of the specific data in the pooled analysis were that in 60 percent of the patients, 42 percent of the identified smokers were advised to quit. 14 percent were given or advised to use nicotine replacement, and 12 percent received referrals.

In other studies looking at narrower populations, for example, a study of patients, smokers with AMI and congestive heart failure and pneumonia, roughly 65 percent had any form of counseling.

CO-CHAIR PINCUS: Are there comments or questions on the issue of gaps?

(No response.)

CO-CHAIR PINCUS: Okay, so we're

ready to vote.

MR. WILLIAMSON: Okay. We will be voting on the performance gap. Please begin voting now.

Okay. We have 12 high, 6 moderate, 0 low and 0 insufficient.

CO-CHAIR PINCUS: Now we move on to evidence, always a sticky one.

MS. HOO: I think with respect to evidence, some of the same issues that we talked about earlier come into play here. In the Committee discussion, I think folks either felt that there was a fair amount of generalizability around the inpatient identification of members and referral for counseling, and identifying that there were opportunities in a patient population relative to teachable moments and such.

But I think again here, some of the challenges were felt around the feasibility, given the complexity of issues that might arise at discharge. But I think in general, the

Committee felt that the evidence was strong for advancing this relative to the available evidence.

CO-CHAIR BRISS: This one strikes me as actually being a little easier in an evidentiary standpoint than the other one, because the therapies have been well-studied and are likely to have less, I'm sorry, the medication therapies are likely to have less generalizability issues than the ones we were talking about in council.

CO-CHAIR PINCUS: I have one question of clarification about the measure itself, with regard to does the information for the numerator have to be in the discharge instructions, and/or transmitted to the next level of care?

MS. LAWLER: This measure really doesn't deal with transmitting to the next level of care. It's strictly that they received a referral. So that could be in the discharge instructions or elsewhere in the chart, where they document that there's a referral to the next

level of care and outpatient counseling, and that they received a prescription, and usually we find those in the discharge instructions as well.

No, it's not required that it be in the discharge instructions. It just has to be documented somewhere in the medical record.

CO-CHAIR PINCUS: I'm just curious about the thinking in the development of this measure, why it wasn't specified that it be in the discharge instructions, the information, because I mean the value of that is that it would be easier, feasibility would be one place to find it. Number two, it would actually be transmitted concretely to the patient.

MS. LAWLER: Give me just a minute. I'm going to look up the data element, because we do give, in our specifications we give data sources where we expect people to find it, what we call a recommended source. Let me just find it here.

So our suggested data sources,

according to our specifications are, and this tells the abstracter where to look. In the discharge summary, the first one listed, a transfer sheet, discharge instruction sheet, nursing discharge notes and a physician order sheet.

So that's where we -- the direction we give people to look specifically for, but we don't say specifically it must be in this particular document. We give them a couple of sources where they can go to look for it.

CO-CHAIR PINCUS: It just occurred to me that it would be in some ways more of an actual -- the effectiveness would be greater if it was explicitly in the discharge instructions, not just simply, you know, wherever you find it in that hierarchy of sources. But anyway, just --

DR. SUSMAN: I just wanted to know in the specification of counseling, was it referral for any form of outpatient tobacco cessation counseling? In other words, there was no requirement about the three components or other

more specific elements? I think that's for the measure developer, please.

MS. LAWLER: So here's our definition for what we require for a referral, and what's acceptable. So we say outpatient counseling may include a proactive telephone counseling, group counseling, individual counseling or e-Health and Internet intervention. Counseling referral may be defined as an appointment made by the health care provider or hospital, either through telephone contact, fax or email.

For quit line referrals, health care provider or hospital can either fax or email quit line referral, or assist the patient in directly calling a quit line prior to discharge. So we want, you know, the patient to get hooked into that system before they leave the hospital, and want the provider to help them to do that. So it can be any variety of types of counseling that I just mentioned here.

DR. SUSMAN: Okay, thank you.

CO-CHAIR PINCUS: Just a question about the evidence. You know, going back to the previous discussions that we had, the extent to which the evidence that's been reviewed reflects specifically the action of referral and prescription from hospitals.

Could you give a kind of a summary of the extent to which the evidence that's been reviewed is specific to that issue?

DR. FIORI: Could you repeat that? I'm sorry. I didn't track it.

CO-CHAIR PINCUS: Okay. So just going back to the discussion we had on this topic before, just that it would be helpful to just have a brief summary of the extent to which the evidence has been reviewed is specific to the issue of hospital referral.

DR. FIORI: So again, I'm sorry, but I can't tell you the proportion of the studies or the number of them that started with the counseling referral at the discharge moment. There are some, but the bulk of them are actually

in other settings.

But in all of the other settings that they've occurred, it's been a very consistent, have a very consistent impact on downstream quit attempts and successful quits.

CO-CHAIR PINCUS: Any other questions with regard to the evidence?

(No response.)

CO-CHAIR PINCUS: I guess we're ready to vote.

MR. WILLIAMSON: We'll be voting on the evidence. Okay, yes. So this is yes, no, insufficient, 1, 2 and 3. Please vote. We're still waiting on -- there we go.

All right. We have 16 yes, 0 no and 3 insufficient evidence.

MS. HOO: In terms of the scientific acceptability, the overall voting was 5 yes on 0 no. On reliability, the scoring was split, with 2 high, 3 medium and 1 low.

Some of the concerns that were expressed here were that the kappa score was very

low, and another comment cited the lack of risk adjustment strategy.

In the documentation, some of the data that was cited related to lack of clarity around the wording, and that was something that was subsequently corrected for or revised in the final specifications. In 131 cases, that had been reviewed, 11 percent or 14 had false positives, and 2 cases had a false negative calculation.

Some of the information that had higher agreement rates were the data element with respect to prescription for tobacco cessation medication, and 64 percent for no FDA-approved tobacco medications at discharge.

CO-CHAIR PINCUS: Comments or questions about reliability? Jeffrey.

DR. SUSMAN: Again, the very, very low kappa this time, I guess, of at least interest if not concern. This is just extraordinary. I assume it wasn't a misprint. I don't know if you have some further comment, or if your

methodologist has some comment.

MS. WATT: Well, Steven, I don't know if you are still on the line, but if you are, feel free to jump in. Again, in this situation, it was an example of where the hospitals were giving themselves more credit basically than we were for particular data elements, and the takeaway message for us was that we needed to strengthen the specifications, and we did that, and I'm sure Nancy will tell us how.

CO-CHAIR BRISS: This one looks like it's like -- so their results were sort of the opposite of those, the results we had in the first one. So very low rates of successful performance. It was like ten percent. So the statistical issues, I think, are going to be like the ones we had, and this one seems to me to be closer to the first one, Tobacco 1 than it is to Tobacco 2, to me.

So it's low kappa, and reasonable numbers, I would say, in the numerator and the denominator, and a low kappa that reflects sort

of a low performance rate.

CO-CHAIR PINCUS: Just a question about the statement that the reason for the low agreement rate was that the patient referral was not always made prior to discharge. That's a requirement?

MS. WATT: Yes, it is. So again, it's not that we want -- we don't want the health care provider to just say well, you need to go call the quit line, or you need to, you know, the next time you go to your physician as an outpatient or, you know, get him to put you into counseling. Can you hear me?

Okay. So we would like the provider to be able to make that referral for the patient before he leaves the hospital. We want to see that that's done before the patient leaves the hospital. And that was the piece that, I think for reliability, really brought this down.

CO-CHAIR PINCUS: So how do you solve that problem? Have you -- I mean if that's the problem with reliability, how do you fix that

going forward?

MS. WATT: It's understanding the requirements. It's getting more familiar with the specifications. I think that in any time that we do testing like this, there's a learning curve, and as they begin to use the specifications, they get more familiar with them, and it's more of a learning, teaching and as we go along, and we find this with all of our measures, even that we use today, when we first begin, the rates are very low, and as people get better at providing the care, the measure rates come up.

I think we'll see the same thing with some of these issues that we see, as people begin to use specifications and don't really understand them, and --

OPERATOR: Mr. Schmaltz has signaled.

MS. WATT: He's our statistician. Steven, did you have a comment?

DR. SCHMALTZ: Can you hear me? Oh, okay. The situation with this measure is more

similar to the previous one, where you had a high rate of false positives. There weren't very many, a relatively smaller number of positives.

CO-CHAIR BRISS: I'm sorry. You think it's closer to Tobacco-1 or Tobacco-2?

DR. SCHMALTZ: Tobacco-2, because there were relatively smaller number of compliance, and these are the false positives.

CO-CHAIR PINCUS: Anyway, Vanita, then -- Tami, you put yours up and that is still up? Okay. Vanita, Lisa, Tami, Caroline, Mady and Jeffrey.

DR. PINDOLIA: When I was looking at the measures and the checklist that has to go on at discharge, if you think about what's going on at discharge, especially now with the competing factor with trying to reduce three-day readmission and you have these complex patients, there's a whole host of things that are being --

These patients, one, are being inundated. So we know that post-discharge, during transition of care, they're not even

getting that critical part unfortunately.

But when you throw in that a nurse or somebody, a pharmacists or someone is going to go through 24 items for a checklist for this item, I think that shows maybe why we're showing such a poor uptake in the reliability, because you can only get through so many when you still have to do all the other competing factors that are being regulated as well. Is there a way to cut that down?

MS. WATT: I'm not quite sure what you're referring to on the 24 item checklist.

DR. PINDOLIA: Oh, I'm sorry. Maybe it's under 2(a), 1.20, calculation algorithm measure logic, and I assume this is what someone is going to go through, to make sure they check that this is the right person for you to have 10 or 12 percent met the criteria.

In their head, somehow they're going to have to make sure they're doing all this stuff, because that's what they're going to be checked against doing chart review; correct?

MS. WATT: I just want to look at the algorithm. Just one moment.

DR. PINDOLIA: Well, it's not -- I don't know. Maybe I'm not looking at the right one. But it's not an algorithm; it's the calculation of the measure logic is what it's listed under.

CO-CHAIR BRISS: I think that this is a little bit different from what somebody would do at discharge.

CO-CHAIR PINCUS: Right.

CO-CHAIR BRISS: They wouldn't at discharge presumably recalculate, refigure the patient age, for example.

DR. PINDOLIA: No, no. No, I understand that. But what I'm saying is there are so many key components of this that need to occur, for you to say you actually did get the right person, and you actually did go through the right process, and then make a referral to a tobacco cessation, which is --

Is it? Okay, all right.

DR. SHEA: Yes. I think I was reading this, but I can't quite find it now. In terms of places where nicotine emplacement is over-the-counter and doesn't require a prescription, the data dictionary states the hospital gets credit for that, even if they don't write a prescription.

MS. LAWLER: Absolutely. It just needs to be documented in the discharge medication list. So, yes.

DR. MARK: It looked like some of the reasons for low reliability was that the physician didn't say why they did not prescribe. So they actually have to write down in the chart "did not prescribe because has." Okay.

CO-CHAIR PINCUS: Mady.

DR. CHALK: A couple of us were commenting here that given the extent to which you respecified and told people, respecified questions, it might be preferable, if you waited a little bit, retested the respecification and came in.

Now my question to Helen, I guess, is this one shot at getting measures done for behavioral health?

DR. BURSTIN: No, it's certainly not. I think our expectations would be we would most definitely be doing another project, likely next year, and actually that's one of the questions I asked Ann, about how quickly, for example, would you learn, even if it's not through formal retesting.

But at least as part of the implementation they shared with us earlier, how soon that could come in. We heard earlier that's towards the end of 2012, certainly beyond this project. But again, I think once that information's available, we'd love to try to get these back in.

DR. CHALK: Right. I think that's a terribly important issue, because one of the things that concerns me about not putting forth a measure is the fact that most hospitals and clinicians aren't going to do anything to get

this right, until they're held accountable. I mean you've got a chicken and egg issue.

But on the other hand, it has to be specified correctly. So I would really encourage you to think about that.

CO-CHAIR PINCUS: Caroline, then Jeffrey.

DR. CARNEY-DOEBBELING: I was going back to the similar issue, of the agreement rate being only 41 percent, and I think we need to focus on the fact that there was a recognition that specifications needed to be made, and specifications were made, with the assumption that that was the reason for the poor agreement rate.

But those new specifications have not yet been tested. So I get hung up on saying that we've made specifications, and we're just going to assume those are okay, so move forward.

DR. SUSMAN: So another variation on this theme is how did the hospitals, in your mind, lead to the false positives? If a

referral wasn't made prior to discharge, where would it have appeared in the abstraction of the hospital records?

I mean did the abstracters just incorrectly abstract data, or what was going on there? It doesn't make a lot of sense to me, the logic here.

MS. LAWLER: Well, I think you're absolutely right. They just incorrectly abstract the data, and our job when we're out there, as reabstracters, is to take that medical record and find that information, and we didn't find it.

So we do sit down with them at the close of our study, and talk about, you know, what we saw and what we didn't see in the medical record, and so that everybody has an understanding that's a time for teaching at that point. But yes.

DR. SUSMAN: Okay, fair enough. I very much second Mady's comment that it seems to me that in this process, at least from my own

perspective, that if there's substantive change to the specifications, at least ideally we would have those when we're considering, particularly when there's such low reliability or kappa as well.

Well, whatever the issue is, if it's going to be a substantive change, that it would be nice to have the data on that change. That's just feedback, I guess.

DR. HANRAHAN: I'm aware of a lot of work that's being done on discharges, and making discharges or discharges from the hospital be more consistent and implemented correctly, and this is all being wrapped into electronic health records.

So I think it's around 30 percent now. Those places that are not doing it are often the behavioral health settings. So, you know, I think that we, this is kind of a procedural measure, in that it's going to prompt this to get into those electronic health records, or get onto the discharge summaries.

And I really think that that's a very favorable thing to do, so it's going to encourage that. These reliability testings, I think is really, you know, it reflects --

What my experience is is that the field is so poorly operationalized around discharges, and it may or may not have anything to do with, you know, tobacco or any particular measure that you might look at.

So I think it's going to do two things. One, it's going to promote a more formal consistency around discharging patients and what you need to cover, and secondly, I forgot.

(Laughter.)

CO-CHAIR PINCUS: Peter.

CO-CHAIR BRISS: I mean essentially we're going to have to choose, I think. You know, in this one, essentially the agreement rates seem to me to be reasonable for cessation meds, and like the other one, are not so great for counseling, because the overall agreement on this one works out to be a little bit better, I

think, because the performance rate is so low, with or without the counseling component.

So this is either like the -- we either decide that it's like the first one and it just passes, or it's like the second one, and we want better operationalization, thank you, for the counseling piece and then review it again.

CO-CHAIR PINCUS: Any further comments on reliability? Okay, Emma.

MS. HOO: I guess I would echo what Nancy said, and you know, coming from a purchaser perspective, I think, you know, I worry a little bit about throwing the baby out with the bath water.

You know, yes, it's not perfect, but directionally, I think we need to drive better documentation of these kinds of issues, and also the infrastructure for capturing the information, that for me, it's acceptable to have lower reliability and recognize that there are refinements that are going to come down the road, that if we don't have a measurement around

an intervention to go with the identification on the back end, I don't think we will have much at the end of the day.

If we defer for a year to re-look at this, I think it just further misses the opportunity to drive better data capture.

CO-CHAIR PINCUS: So are you -- I ask this question. Does that mean you're arguing for an exception?

DR. BURSTIN: There aren't any exceptions on reliability. You just need to make your best guess assessment of what you think the reliability or validity of the measure would be.

CO-CHAIR PINCUS: I guess Jeffrey and then Caroline.

DR. SAMET: So I'm just trying to articulate in my own head the sort of discussion that's going on here. It's that we're almost following -- the sense that I have is that we're almost following the rules so closely that our kind of greater sense that although the

importance is huge, and there's a lot of evidence to say move forward, that a piece of it that isn't quite as good as it should be may put it all on the back burner for another two years, when if we really get it in there and say it ain't great, but it's kind of okay, and that will drive the system, just knowing what the system is to move it forward.

Now did I say that right, because if it is, it leads me to think like, you know, shoot for it's good enough.

CO-CHAIR PINCUS: Caroline and then David.

DR. CARNEY-DOEBBELING: So that's a question that I have. Do we vote on driving the field forward, or do we vote on the measure at hand?

CO-CHAIR PINCUS: David, and then Nancy.

DR. PATING: You know, I feel I'm struggling with this as much as everyone else. Part of me just wants this so badly, but at the

same time, realizing that a poorly designed measure ultimately creates chaos in the field, because you're driving from the outcomes.

Having been there, it's hard for me to support when the measure is not properly conceptualized, and I think that that's what happened with both these. I think Measure 2, the second one, over-reached. It had too many moving parts. It had medication in three components and, you know, it's got to be done on a Sunday versus a Monday. It had too many parts.

This one I feel like doesn't quite go enough. What I want is if you're going to give the pamphlet, why don't you do the brief intervention for gosh sakes, you know, because we know that's what people need. So somewhere in there, if there could be just instructions back for the developer, to go back and take some lessons learned.

I don't want to slow this down, but again being in the field, and having to respond to measures, this would send systems backwards

five years, not to put a good measure out there that does what we really feel needs to happen. That's my two cents on this, is that we really work with them.

DR. HANRAHAN: I really -- point well taken, because we don't want to create more chaos, and as I look at this measure, and the other measures that have come up around tobacco, there is a complexity with them, or shall I say that measures move into very complex environments, and they're going to take on all kinds of different, you know, fussy details as they get implemented.

But I think as they get implemented, my assumption is and one of the questions I asked at the beginning, was the longitudinal effect of what we're doing here, is that it can come back and get its revision.

Now I also think that regarding the reliability description here, that when I read the reliability description there's a lot of flaws in the process because of the lack of or

the tension between, you know, JCAHO and the hospitals wanting to adopt another measure, or people wanting to participate in that process.

In regard to this one, I really think that that's probably one of the reasons why that things didn't go their way. So as a reviewer, I would say this is where I would push the, what is that Helen you said that we can do, we can --

DR. BURSTIN: And ad hoc review in the future to evaluate the measure if it goes forward.

DR. HANRAHAN: Yeah, yeah, and my recommendation would be for all these measures to let, you know, strip them down and get them very incremental, rather than as complex as they are.

It says "Tobacco use treatment provided," or offered. Even that is way too complicated for a system to adopt. So that's what I would say.

CO-CHAIR BRISS: So I was sitting here

thinking about Caroline's question, about do -- how do we evaluate this thing, so I don't think we evaluate based on where we'd like to drive the system. I think you have to evaluate it based on the evidence provided.

Having said that, it looks to me like you could read these evidence. There's some fuzz in -- the medication stuff is reasonable to me. There's some fuzz in the counseling stuff.

The overall performance of the measure, you know, 11 percent false positivity rate doesn't strike me as being out of bounds, right, you know.

So I think a reasonable person could look at these reliability data in total, and decide, you know, passes. Not stellar pass, but pass.

DR. SUSMAN: Well, I worry really about the accountability side. We're looking at measures for both performance improvement, and for Ann and Nancy, if it was just performance improvement, I'd say ehh, good enough.

But when looking at accountability, I worry that we're trying to measure the health system or hospital or a larger aggregate against another, and that the data are substantially flawed as we have them here today, in my mind.

In the end, it's a judgment call. I mean how black is black, how white is white here. I preferably would like to see this come back with the appropriate testing done on the newly-specified measure. I don't think that that has to be a lengthy process, and if NQF has a way to streamline that, fast track it, I think that's great. If we don't, it is what it is.

I would vote that we think further about passing something that could be used for accountability, without having the appropriate rigor.

DR. WEGNER: I've been fast forwarding a couple of years, and thinking about the people that are going to be doing outcomes research, and this is going to be an impossible variable. I mean you're not going to be able to

quantify this, because it's going to be so --

The way it's written, it's going to be so loosely interpreted, the counseling and what happens at discharge. That's completely separate from the issue that was brought up earlier, which is we do have two populations here. We have JCAHO, we have the hospitals, but then we have the providers.

What I'm seeing is more unfunded work, this meaningful use. Has anybody in this room done meaningful use with the problem list and the clinical? That adds a lot, and for a complex patient, and if you have standards of what's the documentation for what you've done, I think you're right.

I think your point about accountability is a big one here, and I think we do need to think about this now, rather than just go ahead and summarily pass it, and then the horse is already out of the barn.

MS. WATT: Could I just remind you that Joint Commission accreditation is

considered to be an accountability function as well?

I think, if I'm understanding correctly, when you're talking accountability, you're talking about use in federal reimbursement programs, and I think that the definition of accountability is broader. We, I'm not intending for this measure to be used as a reimbursement measure.

CO-CHAIR PINCUS: Last word.

PARTICIPANT: It's a really fast one.

CO-CHAIR PINCUS: If you have something absolutely new to say, that hasn't been said by anyone else, that's specifically about reliability and not about feasibility, not about usability.

CO-CHAIR BRISS: So about the reliability, very quickly. I wonder about the implications of the fuzz in the measure that we're talking about. So we're talking about in order to pass the measure, we're talking about somebody has to be prescribed cessation meds,

and prescribed counseling.

What we're worried about is that some people might do something with respect to counseling that's not quite what we would have wanted, right? You know, that's the implication. I wonder if, because they tended to overcall counseling. That's what the disagreement is about, right.

So essentially you're given, you might pass somebody who gives slightly less intensive counseling than what we would really desire, and so I wonder whether that's a big enough difference to make a difference. To be precise, it's over-counting referral for counseling.

(Off record comments.)

MR. WILLIAMSON: Okay. We'll be voting on reliability. You may begin now.

Okay. We have 0 high, 6 moderate, 6 low and 7 insufficient evidence.

PARTICIPANT: All right, everybody. Lunch is available in the back, if you want to --

(Whereupon, at 12:37 p.m., a luncheon recess was taken.)

A F T E R N O O N S E S S I O N

1:07 p.m.

CO-CHAIR BRISS: So I think we'd like to reconvene. If there are people who haven't retaken their seats, we'd like to reconvene.

CO-CHAIR PINCUS: Okay. Let's start reconvening.

CO-CHAIR BRISS: So we want to move on to Tobacco-4, which is the last measure in the series of measures we've been evaluating this morning. We'd like to propose trying desperately to get us closer to on time. We'd like to propose a bit of a streamlined process.

So this measure you'll hear in a second, if you're not real familiar with it, is sort of follow-up after hospitalization. So the series of measures is going to screen, treat and then follow-up, and as a conceptual matter, we think that if you haven't gotten the treatment step in there, it doesn't make much sense to spend a lot of time on the follow-up step.

So we'd like to -- but we want to give

the measure developer feedback, so that they're not blind-sided if there are issues that would come up in the Committee discussion, that could have been found out. We don't want to make them wait a year to find issues that should have been found today.

So we'd like to propose a streamlined process in each of the four areas where we look at the science, we look at the measure performance, we look at the usability and feasibility, and we try to identify for the developers any issues, any additional issues that haven't already come up in one of the first three measures today.

So if the -- so does that make sense? I've got some head-nodding around the table. So I'm going to take head nodding, some head nodding around the table as a sense, and we'll go forward with the streamline process. So could the developers tee up the measure for us please?

Measure 1657

MS. LAWLER: Okay. This is the last

measure in the set. It is assessing tobacco use status after discharge, and so in the denominator we look at all of those patients that again were current tobacco users, and in the numerator, we're looking at patients that were contacted between 15 and 30 days after hospital discharge, and that the information regarding the tobacco use status is collected.

Actually, there are three data elements that we want information collected on, and that's whether or not the patient is attending the referred counseling; whether or not they're taking the medication; and then whether the tobacco use status at that point in time at which they are contacted.

CO-CHAIR BRISS: And I'm also pinch-hitting for -- I wasn't, this wasn't one of my assigned measures, that I'm pinch-hitting for, whoever was going to report out on this one. So I will try to reconstruct where we were on this one.

So on the evidence side, the work group

generally, consistent with everything else we've done in tobacco, found a high preventable impact, identified an existing performance gap, thought that the evidence decision logic was generally high or moderate, and seemed relatively comfortable with the body of stuff on importance to measure and report.

So are there -- in this section, does anybody have additional kind of comments related to this measure that we haven't already made on the first three? Harold.

CO-CHAIR PINCUS: I guess my biggest concern with this measure, this actually came up at the MAP Steering Committee, is this places a fairly substantial burden on hospitals, to follow up 30 days later, and it's unclear what the benefit is, because the hospital would just be finding out whether or not there was any follow-up.

There would be no additional counseling or other information, and it just seems like an odd kind of situation, where one

can imagine the hospital sort of hiring college students, to call people and say, you know, did you ever follow up with counseling? And if the person says "no," what happens next?

If this was a measure for an ACO or for a medical home, you'd think about it totally differently, where they have some responsibility to follow up and also to do something about it. But it seems like a great effort and expense to simply document something, without being able to improve the situation.

DR. CHALK: On the other hand, Montefiore Hospital in New York has taken exactly the opposite position, which is that since they know that the majority of their return readmissions one week following discharge from even a medical situation in their hospital, are connected either with substance use or psychosis, and not surprising that it's an urban hospital, they've created --

CO-CHAIR PINCUS: They also are an ACO that is at risk.

DR. CHALK: Yeah, but that isn't out of -- now they are, but that isn't out of which they created their community care, what would you call it, their unit, community care unit, that does this very thing, that follows up between a week and 30 days to track people and to see whether they're doing what they're supposed to be doing with their treatment.

I guess I'm not -- not something to dispute, but I think that that makes a lot of sense for them to do that, but they have to have a business model that allows for that. Again, it's not just tracking; it's tracking and intervening, and this --

(Off record comments.)

CO-CHAIR BRISS: So I was also -- I think -- oh, I'm sorry. Karlene, you want to go?

MS. PHILLIPS: I'd just like to echo what you said. We have a process on my behavioral health unit, where we call patients within three days of discharge. We actually only reach probably 30 percent of the patients

who we discharge. The rest of them we cannot reach by phone, and they do not respond to letters, those kind of things.

So tracking this kind of information I think would be extremely difficult, and it is very time-intensive to make all those phone calls to patients who discharge.

CO-CHAIR BRISS: So I was another person on this one who was a bit more sympathetic to the measure and the intent of the measure. So I think that it's quite possible that some hospitals might find that when they referred folks for counseling, that very low proportions of people actually received the counseling that they were supposed to get.

It seems to me that that would be actionable information, that you wouldn't have an easy way of knowing otherwise, and in a world where we're trying to do better about coordinating across settings and contexts, and in a world where we know that the current state of the art on handoffs between settings is

generally, I'll say politely, leaves something to be desired, even when everybody agrees that it's critical, right, that we have some room to move. This might provide some impetus to move in an appropriate direction.

DR. FIORI: On behalf of the Technical Advisory Panel, I'd just like to share some of the scientific rationale for why this was recommended.

I mean sort of a theme of the discussion that we've had all morning, and that is that tobacco addiction is unique in terms of its morbidity burden on the health of our patients, and that we clearly are failing currently to maximally utilize health care encounters, to ensure that patients are more likely to leave those health care encounters with evidence-based treatment. So at the core, that's what drove it.

But in addition, there are data in the tobacco cessation literature, and one of the rationales for having the fifth of the 5(a) being

arranged follow-up, that by the simple act of arranging a follow-up contact, that smokers are more likely to make a good attempt and even to quit.

That's not a robust database, but there are data that support that, and it's because of that this notion that follow-up contact is associated with patients more likely engaging in the cessation treatment that was prescribed. We had top three at discharge.

It also provides an opportunity to measure outcome data on this critical variable, and that is another important potential.

It also is consistent with our evolving comprehensive care model of disease in general, where more and more, whether it's post-delivery of a baby, congestive heart failure, diabetes, management, there is a post-discharge follow-up. So it's consistent with that.

This might be a stretch, but I'm going to mention it anyways. You know, there's such

an increasing evidence on hospitals to decrease their 30-day readmission rate, and we know if they quit smoking, they are less likely to return to hospital.

I think it has the potential at least, and I acknowledge this is a bit of a stretch, but has the potential to decrease 30-day readmission rates if patients, particularly those with COPD diagnoses, pneumonia, all of the respiratory diagnoses, go home and stay smoke free post-discharge.

So those were the rationale that was used by the Technical Advisory Panel, to say that understanding the incredible burden and ma'am, I'm sorry I don't know your name, but Karlene, that it is a burden.

Following up patients in general is very difficult. But just the powerful influence on helping patients to quit, we felt, warranted doing it in this instance.

DR. SHEA: One question was, in following up with that, is that the

specifications seem that you only get credit for actually making the contact. So that also means many attempts to try to get the contact, you know, if you're going to try to do well on the measure side.

I was wondering if there was credit for the attempt, versus actually getting the contact with the person.

MS. LAWLER: No, there's not credit for the attempt. But we did make revisions to the measure. I'm sure that you must have the final version of it, because I see some changes up here, just in the description.

But there are some exclusions to the context that may help, you know. You're not going to contact people that are discharged to another hospital for care. People who are not in the United States, we found that in areas where people go for vacationing a lot. So you're not going to follow up with those people.

People who are discharged to jail perhaps, prison, we're not going to call them.

We did make allowances for those lost to follow-up. So in other words, if you take a number, have a number of attempts and you can't get in touch with the person, then we consider them lost to follow-up, and we send that to what we call a Category B, which is not in measure population.

So there's, we tried to work in a number of situations that would deal with those kinds of issues.

DR. EINZIG: Just a comment to follow up on what you were saying also. It's a question of who is the best person to be calling the patient or family. If it's the hospital calling, who has no established relationship versus a primary care provider, somebody who has an ongoing quality relationship, I think that may make a bigger impact, a bigger difference.

Thinking of it from a patient perspective, if I were discharged from the hospital, I wouldn't want to remember that event. If I got a caller ID and saw the hospital

calling me, the first thought I might have would be okay, I have to pay the money or what's going on.

So just thinking in those types of terms, not just from a hospital perspective, but also a patient perspective.

CO-CHAIR BRISS: So I don't think anybody has raised yet the issue of -- so on the second of these measures, we talked a lot about how the intervention was defined, and most of the evidence is about the month's worth of follow-up.

So I thought that the intent of this measure was in part to line up what you were recommending people to do for counseling, with the level of follow-up that has the best evidentiary basis. So that struck me as positive in the set of measures taken as a whole. You might be able to play that up a bit more than you did in the set of measures, sort of as you're bringing them back.

DR. BURSTIN: And just one other

thought. I don't want people to think that just because the measure is complex, we shouldn't do it. I mean it's difficult collective.

If it's really important, it should be done. In some ways I find this somewhat analogous to the fact that in fact, surgical site infections required 30 days of surveillance.

So I think as you're making the case coming back, I think it would be helpful to in fact emphasize that in fact the full episode after hospital care, and referring back to that Cochrane review, was in fact the in-hospital intervention plus 30 days.

So I think actually being able to see at the end of the day whether you're successful or not actually has some face validity for me at least, in terms of a similarity to what you would need to do for surveillance or on SSIs, and if smoking is so important, you know, I think there's a way to play that better perhaps.

CO-CHAIR BRISS: So clearly this measure is going to have a number of feasibility

issues that you're going to want to make a case for. We've raised a few other issues. Does anybody else want to raise an issue that hasn't been raised already, in terms of importance to measure and report on this measure?

CO-CHAIR PINCUS: Just the only thing I would add, and this also came up at the MAP, is it's not just smoking, but there are many things that ought to be followed up with post-hospital, and as the Joint Commission considers sort of what kind of package of follow-up activities are sort of considered essential or important as a measure, to sort of somehow integrate it so that there's not a lot of little different measures.

CO-CHAIR BRISS: Is there anybody else -- I'm sorry, Jeff.

DR. SUSMAN: Just briefly. It would be pretty easy to know what the outcome is, are they smoking or not. It seems to me that's really what we're trying to drive, is the ultimate outcome of not smoking.

While obviously it's a very short time period, if you've been effective at the hospital at providing the brief counseling, the prescription, it would be sort of interesting that causal pathway to outcome, of whether they're smoking or not.

CO-CHAIR BRISS: Anybody else, things that haven't been raised already in importance to measure and report. I'm sorry, yes.

DR. PATING: So with regards to we've found Questions 2 or Tobacco-2 and 3 not to be valid measures, what are we actually measuring with number four, assessing the status, whether they're not smoking?

I guess what you've done is you've provided smoking screening. There's been no mandate for any intervention, and then we're measuring some hypothetical kind of non-outcome.

So it would be up to hospitals to do something in the middle there. Is that kind of what we're, you know, the process of you think

of these two as linked indicators? Isn't 1 and 4 together?

CO-CHAIR BRISS: I was thinking more like three and four together. So three essentially says thou shalt give, thou shalt refer for or provide both medications and counseling, and this essentially measures whether people got those, and whether they're smoking.

DR. PATING: Right. So there would be a backwards implication, because we've helped them with alerts to 2 and 3. They're no longer on the table. But there would be an implication, I think, by passing 4 of expecting something.

CO-CHAIR BRISS: Two and three are going to come back, pending additional reliability testing.

DR. PATING: I see.

CO-CHAIR BRISS: So what I'm anticipating happens going forward is that the Joint Commission's going to essentially do one

more feedback loop. They saw originally that the reliability testing was low on Measures 2 and 3. They've fiddled with the guidance that they've given people to try to improve that.

They'll add another feedback loop that in July improves reliability testing, and then they'll bring those back in conjunction with this one.

DR. FIORI: But Peter, to be specific, there's no implication that 4 would be endorsed today, in the absence of 2 and 3, and I think that's what the point was.

CO-CHAIR BRISS: Which is why we're doing a streamlined process right now on 4 and not voting.

DR. PATING: Not voting, yeah. I think that was --

(Off record comments.)

CO-CHAIR BRISS: So anybody else on -- yes.

DR. WEGNER: If we're giving you advice, you have another group to put in your

denominator, and those are people who do not speak English.

DR. FIORI: Yet if I could add to the test group rural hospitals, where there are great distances to follow up, particularly mental health rural hospitals, where mental health patients have traveled long distances.

CO-CHAIR BRISS: So anything else on important to measure that hasn't been raised already?

DR. GOPLERUD: Is there any advice about kind of two measures in one or three measures in one, where this measure came initially as a process measure? Did you do a contact, a follow-up contact?

That was a process measure, in order to get to an outcome measure, and there was a similar measure or set of measures that were approved by a behavioral health committee around measuring depression level within six weeks following initiation of treatment.

There was a process measure which was

did you do a screen, and then an outcomes measure, which is "and what happened."

Is there either a positive with that, or how would NQF give advice about that kind of a complex measure?

DR. BURSTIN: Yeah, and Jeff actually chaired that committee, as I recall, so he probably could speak better. But I think the idea was essentially just in a broad term measurement-wise.

The idea was that the process measure just meant you actually gave somebody the assessment tool to complete, which was required to do the measure, and the n the measure itself was the delta of the PHQ-9 over a six-month period.

So I mean and those were paired, right. Anything you want to add, Jeffrey? You know this better than I do.

DR. SUSMAN: Well, the only thing I'd add is sort of the broader issue of where we should be driving to, which is the outcome, and

in some ways, all these process measures that we're putting into place I think sort of is we're missing the forest for the trees, you know.

We start specifying every step along the way, and what if the real follow-up was hospitals have to be accountable for their patients longitudinally in the outcomes for smoking, as much as anything else they do. I mean we have it for all sorts of bizarre things that occur in hospitals, to a relatively small number of patients.

But we ignore the fact that smoking probably is the number one cause of preventable morbidity and mortality. So you know, my soap box would be let's get you involved with actually getting to the outcome of smoking cessation over the long run.

CO-CHAIR BRISS: As you're a communication consultant, I don't recommend that we pitch to the Joint Commission that way. We're trying to add an additional crazy thing that you should be responsible for. Yeah.

So last comment on this point, and then I really want to move us on.

DR. PINDOLIA: Just to echo what Lisa said, but I just want to see if there's anything Joint Commission can do to segregate maybe just two types of hospital types, because if you look at the hospitals in the downtown area, I know in the Henry Ford Health System we have five hospitals.

In one hospital, 50 percent of the patients have no telephones for us to ever reach them. But the other four, we would have a huge success rates.

For them to be compared side by side with physician groups, to say who I should admit to and all of that that where it's leading to, that's just something to look into, if there's a way to segregate the two types.

DR. CARNEY-DOEBBELING: It will be the safety net versus non-safety net hospitals, which may be a way to separate that.

CO-CHAIR BRISS: So I want to move us

on to reliability testing. Will whoever's controlling the screen scroll us down to the results of reliability testing please?

(Pause.)

CO-CHAIR BRISS: So these were actually better to me. Two percent false positives, five percent false negatives, .7, greater than .7 kappa. So and the work group generally called reliability, actually both reliability and validity moderate.

So does anybody have additional comments for the developers on reliability or validity, that haven't already been raised on at least one of the measures this morning?

(No response.)

CO-CHAIR BRISS: Hearing none, I'd like to move us to usability. In general, the work group thought it was moderate. So in the work group, we had the same theme about the difficulties in reaching people post-hospitalization. So that's come up several times already today. I don't think we

need to hear that again. Does anybody else have usability issues that haven't already been raised?

(No response.)

CO-CHAIR BRISS: Hearing none, and we've already raised a number of feasibility issues with this measure. Is there any other feasibility advice you'd like to give to the developer on this measure, that hasn't already been raised?

(No response.)

CO-CHAIR BRISS: Hearing none, is there anything else that the developer is desperate to hear from us, before we close --

MS. LAWLER: Yes. We want to hear yes.

(Laughter.)

CO-CHAIR BRISS: That may take a little longer, as a matter of fact. Okay. Then we will close the discussion on this one and move us on.

MS. FRANKLIN: So that brings us to

Measure No. 0028. The measure developer is AMA/PCPI, and I understand there will be some folks on the call who will also discuss this measure.

We also have Sam Tierney in the room. This is the measure, Preventive Care and Screening, Tobacco Use, Screening and Cessation Intervention, and we'll have the developer tee it up for us.

Measure 0028

MS. TIERNEY: Good afternoon, everyone. Thank you. Thank you for your time.

So since you've already reviewed a number of measures related to tobacco use, screening and intervention at the inpatient facility level, I won't restate the importance of the intervention and the well-established benefits of tobacco cessation interventions.

I'll instead focus my comments on just some of the key features of the measure, Measure 0028 that you have before you, including its history and current use. This measure was

designed for use in the ambulatory setting, to assess clinician performance and ultimately improve quality.

Given that data indicate that approximately 76 percent of current smokers have at least one outpatient office visit each year, there is a significant opportunity for the clinician to screen (off mic) and deliver effective cessation interventions.

This measure was originally developed in 2003, with significant update in 2008. It was developed through the consensus of a multi-disciplinary, cross-specialty expert work group that was convened by the AMA, Convened Physician Consortium for Performance Improvement, as part of a set of performance measures related to preventive care and screening services.

As originally developed by the work group and endorsed by the NQF, the single measure presented for your review today existed actually as a pair of two separate measures, one measure

focusing on screening, and a companion measure focusing on brief intervention, specifically advising smokers to quit.

The updated version of the measure combines the essential elements of the original measure into one measure, screening and cessation interventions for patients identified as tobacco users.

According to the guidelines from the Public Health Service and as discussed earlier, while screening alone increases the rate at which clinicians intervene with their patients who smoke, it does not by itself produce significantly higher rates of smoking cessation.

So cessation interventions are also required to impact the outcome and interest. As a result, the work group that developed this measure agreed that an enhancement to the previous version of the measure would include both components as part of one measure.

The original version of the measure

has been utilized in a number of national programs, including CMS' Physician Quality Reporting initiative or system, as it's now called, and as a core measure for Stage 1 of meaningful use.

The updated measure was used in PQRS in 2011 and is currently in use in PQRS 2012, and it has also been proposed as a core measure for Stage 2 of meaningful use. So that's a high level overview of the measure. Thank you.

CO-CHAIR BRISS: And for the Million Hearts Initiative.

MS. FRANKLIN: Is there -- Bernadette.

DR. MELNYK: We just received the update last week, so our Subcommittee really didn't have a chance to get together and meet. But Vanita

and I just met over lunch, so we will bring comments, at least, from that particular discussion.

So there's no doubt about it, in terms

of the evidence for high impact, we feel.

CO-CHAIR BRISS: Any discussion?

(No response.)

CO-CHAIR BRISS: Hearing none, so shall we vote?

MR. WILLIAMSON: We'll now be voting on impact. Please begin voting now.

Okay. 19 high, 0 moderate, 0 low and 0 insufficient.

MS. FRANKLIN: So Bernadette, could you --

DR. MELNYK: Sure. So in terms of the performance gap, there is a variation that exists. There are suboptimal rates of asking and advising to quit, as well as prescribing pharmacotherapy.

MS. FRANKLIN: Is there any discussion from the work group members on this one? Vanita had a comment.

DR. PINDOLIA: So when Bernadette and I were discussing this in our brief discussion, one comment that I had was as noted in there,

their end goal, this is an intermediate measure, end goal is to have successfully quite smoking, to decrease the heart rate and get the actual outcomes, on heart attacks get the actual outcomes.

My question to the developer, since this was developed in 2003, there's measurements in 2008-2009 which shows the gap in the counseling and the increase, has there been any attempt to looking at the actual end goal?

My question, I guess, and I asked Helen earlier, is at what point does NQF say an intermediate measure needs to be measured to see if it actually met its overall end outcome goal? That was my question to the developer.

MS. TIERNEY: So thank you. I think that that is ultimately where we'd want to get to. I think that the data from the national landscape, as well as the data from PQRS, indicates that there's still quite an opportunity for improvement here related to the process measure.

So I think we wouldn't want to necessarily do away with that until maybe rates were much higher than they are right now.

I think that we could consider, and probably will consider the next time we convene the preventive care and screening work group that developed this measure, to try to develop possibly an outcome measure.

But I think we certainly see value in a process measure, when we know that the rates of adherence to such a process measure are poor.

CO-CHAIR BRISS: And it has to be through the cessation therapy. If that's not prescribed, it can't help anybody, right?

DR. PINDOLIA: Well, I guess that was part of my problem too, with the three minute counseling session and what the impact of that was.

So it's kind of -- to try to say if we continue this for five years now, or four years, continue for another one or two or three years, are we advocating the right way to counsel and

move this forward? So it was trying to see if there's data.

CO-CHAIR BRISS: Any other comments on this one, on this section?

(No response.)

CO-CHAIR BRISS: So can we vote?

MR. WILLIAMSON: We'll be voting on the performance gap. You can begin voting now.

Okay. We have 13 high, 6 moderate, 0 low and 0 insufficient.

MS. FRANKLIN: Thanks. So moving on to 1(c), evidence.

DR. MELNYK: The evidence shows that a meta-analysis was conducted, that really showed that free physician advice significantly increases long-term smoking abstinence rates. We had a couple of questions. One, how long term is long term.

CO-CHAIR BRISS: So Dr. Fiori, would you like to answer?

DR. FIORI: For the meta-analysis. Oh, I'm sorry. I'm sorry.

MS. TIERNEY: So Dr. Fiori knows, I'm sure, much better than I could even explain. But so we did base the measure off of the United States Preventive Services Task Force recommendation, which is also based on the Public Health Service's guideline.

I'm not sure that they got to that level of specificity in the description of the evidence report. So I actually I don't know if Dr. Fiori has anything further to add.

DR. MELNYK: We were also wondering how many subjects were included in this meta-analysis, because the question is about statistical power. As a clinician, particularly in terms of working with people on behavioral change interventions, we know that it often takes multiple sessions.

This gets back to the long-term outcome, you know, how long is long term? If there are tons of subjects in this meta-analysis, did we pick up statistically significant difference, because the power was so

great.

First is the clinical meaningfulness of this. So these are just questions that Vanita and I had.

CO-CHAIR BRISS: So we can turn again to AMA, and ask AMA to quote Dr. Fiori.

MS. TIERNEY: I do have a document in front of me, and --

CO-CHAIR BRISS: Or you can feel free to turn to Dr. Fiori, if you would like to.

MS. TIERNEY: Okay, if he wouldn't mind speaking to this.

CO-CHAIR BRISS: Yeah.

DR. FIORI: It's kind of like a cone of silence.

(Laughter.)

DR. FIORI: To the question of long term, the criteria for inclusion for meta-analyses in the 2008 and prior Public Health Service guidelines was at least six months post-quit date.

The reason that that date is taken is

that the bulk of relapsing occurs in fact more than 50 percent occurs within the first two weeks and by six months, you're at above 80 to 90 percent of relapsing had occurred. So people declare themselves within six months.

There clearly are individuals who relapse later, but the bulk of them relapse by six months, and that's why that was the criteria. This recommendation is very much a clinic-wide recommendation.

The notion that if every physician does a little bit, that shows even a small increase in quit rates, the clinic-wide impact of that is going to be enormous.

So it was mentioned earlier, I think by Caroline, that the data shows strongly that the more counseling you do, the higher the quit rates. But even brief counseling, particularly by physicians, is effective, and the study included seven -- the meta-analysis included seven studies.

Ma'am, I'm sorry, but I don't know the

sample size. But most of them were on the larger size, at least 300 or more participants. I think that covered the questions.

MS. FRANKLIN: Is there any other comments on 1(c)?

DR. SAMET: Just a question. I'm not sure it's timed at the right time. So this is for doing the screening. Is there a frequency? I mean this is in the outpatient setting. So when we're talking inpatient, it was an uncommon event.

So when it happened, it wasn't an unreasonable thing. Okay. So but in the outpatient setting, where it's a common event, are we talking about once a decade, once a year?

(Off-mic comment.)

DR. SAMET: Every two years, okay. I'm sorry, thanks.

DR. BURSTIN: And I was curious if the developer could actually respond to that two-year window as evidence.

DR. SAMET: But I'd just note there,

is there data behind the two years, or was it picked out of the sky?

MS. TIERNEY: Yes, that's a good question. So I think that in selecting two years, the group was trying to be sensitive to burden issues, not wanting to necessarily have to have, since many people don't smoke, having to ask a long-time non-smoker repeatedly whether they still smoke or do not smoke.

Also, I think, you know, the fact that it's a two-year time window doesn't preclude someone from asking more frequently. It's just trying to set sort of a minimum standard. So that was -- and also, as I said, it was part of a suite of measures related to preventive care and screening. So many of the measures take place over a two-year time window.

DR. SAMET: But from what I hear, it made sense to someone. It wasn't data-based.

MS. TIERNEY: No.

DR. CARNEY-DOEBBELING: A question about the measure itself that I wanted to be

clear on. In the two-year period, is it the same provider has been engaged with that patient for two years, same provider group, same group working on EMR?

How is that looked at if a -- especially in the case of someone with chronic disease using specialists. They may not be accessing a single provider twice in that period of time.

MS. TIERNEY: Yes. So it is supposed to be the single provider, and in fact, the reason that we added language in the denominator about making sure that the patient was seen twice for any visit, was to ensure that that patient was under the regular care of that clinician.

So that was the intent, and it's at least in the claims system, it would be assessed through the use of any of the CPT service codes that are associated with office visits. I'm not sure about the EHR. I do have my colleagues on the phone, so I don't know --

DR. CARNEY-DOEBBELING: The

denominator statement just says all patients aged 18 years and older, who were seen twice for any visits, or who have at least one preventive care visit through a two-year measurement period. So it doesn't specify with a single, with the same provider.

MS. TIERNEY: Right. But that is the end time. I can appreciate your question, though.

CO-CHAIR BRISS: Any other questions or comments? (No response.)

MR. WILLIAMSON: Okay. We will be voting on the evidence. Again, a reminder. This is a yes, no, insufficient question. So 1 is yes, 2 is no and 3 is insufficient. You can begin voting now.

Okay. We have 17 yes, 1 no and 1 insufficient.

CO-CHAIR BRISS: Our favorite topic for today, reliability and validity.

DR. MELNYK: Reliability, there is evidence to support high and stable reliability.

The reliability at the average number of quality reporting events was stable, in the .86 to .88 range.

MS. FRANKLIN: Any other comments on reliability from the work group? Or the remaining steering committee?

(No response.)

MR. WILLIAMSON: Okay. We will be voting on the reliability.

MS. FRANKLIN: We have a comment.

DR. MARK: Sorry if I missed this. So when you use this measure, do you limit it to physicians or providers who have a given number of patients, or a given number of --

MS. TIERNEY: So no. The measure is used in a number of different programs, and in the PQRS program, I don't believe there is a limited number of eligible cases. Did somebody -- I'm sorry. I think I heard somebody trying to speak on the phone.

DR. NAEGLE: (breaking up) There is not a minimum number of patients for these

measures to be applicable.

DR. MARK: Okay, because I was just noting that the reliability varies a lot, depending on the minimal number of cases that you have per provider, if I'm reading this right.

CO-CHAIR BRISS: So any other questions or comments?

(No response.)

CO-CHAIR BRISS: Hearing none.

MR. WILLIAMSON: All right. We will be voting on the reliability. You may begin voting now.

Okay. We have 8 high, 11 moderate, 0 low and 0 insufficient.

CO-CHAIR BRISS: So that brings us to validity.

DR. MELNYK: An expert panel of 30 supported face validity. Content validity was established as well.

CO-CHAIR BRISS: So a mean of four plus on a five-point scale. Questions, comments or concerns? Yes.

CO-CHAIR PINCUS: Just is there any evidence beyond threat, beyond face validity, in terms of any evidence that was put together?

MS. TIERNEY: We don't have anything beyond the face validity.

CO-CHAIR BRISS: Questions, comments, concerns?

(No response.)

CO-CHAIR BRISS: Hearing none, shall we vote?

MR. WILLIAMSON: We will now vote on validity. You may begin voting now. For validity, we have 6 high, 11 moderate and 2 insufficient evidence.

CO-CHAIR BRISS: Okay. So that moves us to usability. So to usability, please.

DR. MELNYK: Usability on this measure, I believe it is high.

CO-CHAIR BRISS: Questions, comments, concerns? Yes.

DR. PATING: Can I just -- the only data elements that I can see then where you did

a three, or I believe it's a ten minute best practice-driven intervention; is that correct? That's the major numerator, the data collection?

MS. TIERNEY: Yes. So the numerator talks about a cessation intervention, which is then later defined as either brief counseling, three minutes or less, and/or pharmacotherapy. So it is quite broad.

CO-CHAIR BRISS: Other questions or comments or concerns?

CO-CHAIR PINCUS: So do you have any sort of qualitative feedback about the experience of the clinicians being assessed and sort of their sort of qualitative perception of the usability and how it's being used in their, and some of the issues that they've encountered?

MS. TIERNEY: That's a good question. So occasionally we get comments and questions about, related to the use of the measure in the PQRS program. I'm not sure Kendra, if you have any insights with the -- in that regard.

I think generally, we received

positive feedback. We know that many of our work group members have implemented this within their practices, and haven't received any specific negative feedback.

I think part of the reason it's so broad is that, you know, it really allows for the clinician to determine, on an individual patient basis, what might be appropriate.

CO-CHAIR BRISS: Yes.

DR. SUSMAN: Certainly, there are whole communities of physicians who are using this measure, or one extremely similar. So I think that there's pretty widespread actually ramp-up, at least, in select areas.

CO-CHAIR BRISS: Yeah. This one's being used in meaningful use and PQRS and in God knows how many other places. So if this one can't pass a usability test, I'm not sure what could possibly pass. So votes.

MR. WILLIAMSON: We will now vote on the usability. You can begin voting now. We have 15 high, 3 moderate and 1 low.

CO-CHAIR BRISS: So the last one is feasibility?

DR. MELNYK: And feasibility is good. It will become even better as more primary care practices incorporate electronic health records.

CO-CHAIR BRISS: Questions or comments?

(No response.)

CO-CHAIR BRISS: Hearing none, let's vote.

MR. WILLIAMSON: We will now vote on feasibility. You may begin voting now. We have 12 high and seven moderate.

CO-CHAIR BRISS: Any final comments before the overall vote?

DR. SAMET: We can make an overall vote. Is there a time in this where we, as this moves forward, that we give some comments about to things to be reflected on? I'm still bothered by the two year, based on no data time frame, which will need to be done.

CO-CHAIR BRISS: Why don't we take, unless somebody objects, why don't we take now to do that? Is that -- do you want to make any other comments besides expressing the concern?

If we're going to express that concern, you know, it seems to me that from an evidentiary standpoint, almost every periodicity is of any kind of preventive screening or testing is to some extent arbitrary, right.

They're getting to be some cost-effectiveness counter-examples to that. But there aren't huge numbers of those, right?

DR. SAMET: I would say not arbitrary, but I would say in need of a lot more understanding. I mean because it doesn't have to be arbitrary. That's data that could be out there, that one could look at, and people just haven't looked at it.

CO-CHAIR BRISS: So any other comments?

DR. SUSMAN: Yeah. I guess I would

just -- I believe there are data about what percentage of people begin smoking, stop smoking, and could go then to Jeff's question. I can't cite that offhand, but in a prior life, I did some of that work.

CO-CHAIR BRISS: So anybody else, comments before we do a final vote?

DR. BURSTIN: It would certainly be helpful, again, I don't know what the evidence here is, but certainly the PCPI is relying on the USPSTF, and I know, having overseen that process, how difficult it is to get at periodicity, and I don't know whether that's been updated in any of the updated recommendations. But that would certainly be a place where it should come from, would be what the USPSTF says the evidence or the guidance says.

I don't know if the evidence, you know, of two years is reasonable. It seems like a long time to me as a clinician, but I could think the only other second point I'd raise is if the

measure does include CPT-II codes for both three minutes and then three to ten minutes of testing, three to ten minutes of counseling.

Just in a conversation with Vanita earlier, it certainly sounds like it might be useful, as we think about the world of gathering data for comparative effectiveness, to actually be able to stratify the results. Whether it was less than three or three to ten, it should begin getting to some outcomes, and see if in fact we can gain some knowledge out of having this measurement in place.

CO-CHAIR BRISS: Dr. Fiori.

DR. FIORI: Well, two things. The 2008 guideline recommends that screening take place -- the 2008 Public Health Service guideline recommends that screening take place at every visit for every patient. So pure and simple, and to the issue of time counseling and outcomes.

As was mentioned earlier, there's a clear dose response relationship between three

minutes or less, three to ten and greater than ten. But even the minimum boosted success rates.

CO-CHAIR BRISS: Isn't it true that there are, there's a lots of pushback in the clinician community about having to ask your 60 year-old lady, lifelong non-smoker about tobacco at every visit, right?

DR. FIORI: And I think what's happened in practice over the last 20 years is that this really has been part of the vital signs, often collected during the assessment by the medical assistant or roamer. There's actually very little pushback from it by patients any longer.

CO-CHAIR BRISS: Anybody else, comments or concerns?

(No response.)

CO-CHAIR BRISS: So why don't we move to the overall vote?

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement. This

is a 1 for yes and 2 for no. You may begin voting now. We're still waiting on one response. There we go. Unanimous approval, 19 yes, 0 no.

CO-CHAIR BRISS: So it's nice to finally have an easy tobacco one.

Measure 0027

MS. FRANKLIN: Our next measure is 0027, Smoking Cessation Medical Assistance, and it's three parts. Advising smokers to quit, (b), discussing smoking cessation on occasions, and (c), discussing smoking cessation strategies.

Our lead discussant for that is Dr. Lynn Wegner. The measure developer is NCQA. If they could tee up the measure for us.

MS. ALAYON: Hello everyone. I'd like to introduce myself again. This is Dawn Alayon. I have here my colleague Mary Barton, and we'll be here to present on this measure. This smoking measure has been part of NCQA's HEDIS set since the late 1990's.

It's a survey space measure which is

administered through the CAHPS survey for the Medicare, Medicaid and commercial product lines. It's a population-based measure, where it's a self report. We are trying to capture how health plan members are getting their smoking cessation and tobacco use cessation advice.

So as noted, there are three indicators that Angela -- so advising smokers and tobacco users to quit, discussing cessation medications and discussing cessation strategies.

This measure aligns with the USPSTF guidelines. When it was originally endorsed by NQF, it did not have the tobacco use as part of the measure and back in 2008, this measure went through reevaluation. The measure went through cognitive testing, in addition to face validity.

This data collection is through the health plans. It's done through a rolling average methodology. So we look at two consecutive years' worth of data to reduce the

health plan members' burden to capture this data. This measure is done through the State of Health Care Qualities, Quality Compass and America's Best Health Plans, and it is selected for meaningful use, Phase 1.

CO-CHAIR BRISS: And the discussant is Lynn Wegner.

DR. WEGNER: I'm actually going to ask someone else on the work group to present this. I was not able to be on the conference call, due to a scheduling conflict.

CO-CHAIR BRISS: Would anybody else like to volunteer?

DR. EINZIG: So this is looking at advising smokers to quit and offering recommendations to quit, and offering medication options, looking at adults 18 and over.

This is process. In terms of importance of the study, I think this is a fairly straightforward study also, so I'm not sure what else there is to add, other than --

CO-CHAIR BRISS: So would anybody like to say anything further about tobacco being an important issue before we vote?

(Laughter.)

CO-CHAIR BRISS: So let's vote.

MR. WILLIAMSON: We'll be voting on the impact. Again, this is 1 high, 2 moderate, 3 low and 4 insufficient. You may begin voting now. We need 18, then. We're missing one person.

CO-CHAIR BRISS: Could you revote? We're missing one.

MR. WILLIAMSON: There we go. All right. Unanimous. 18, 0, 0 and 0.

DR. EINZIG: In terms of performance gap, I don't think there's really much else to add there, other than we could do better. If anyone else has any comments?

CO-CHAIR BRISS: So it looks like the typical levels are -- the mean is like something around 50 percent, is that right?

PARTICIPANT: 75 percent.

CO-CHAIR BRISS: 75 percent.

(Off record comments.)

CO-CHAIR BRISS: So questions or comments or concerns before we vote?

(No response.)

CO-CHAIR BRISS: Hearing none.

MR. WILLIAMSON: We'll now vote on the performance gap. You can begin voting now. All right. We have 12 high, 6 moderate, 0 low and 0 insufficient.

CO-CHAIR BRISS: So quality, quantity, consistency of the science?

DR. EINZIG: So looking at the evidence, it looks like they extrapolated data from USPSTF. I think we've all agreed that there's lots of studies that go behind that. I don't believe that there was anything else mentioned beyond that in the paper.

CO-CHAIR BRISS: So we've talked about the effectiveness of education a number of times this morning. So anybody have anything else to add that hasn't already been said on the

evidence for this one?

(No response.)

CO-CHAIR BRISS: Hearing none, let's vote.

MR. WILLIAMSON: We will now vote on the evidence. Again, this is a yes, no, insufficient question. You may begin now.

CO-CHAIR BRISS: David, you're showing record-breaking timing. You should keep it up. Good for you.

MR. WILLIAMSON: We're still missing two responses.

MS. FRANKLIN: If everyone could revote one more time.

MR. WILLIAMSON: There we go. All right. For evidence, 18 yes, 1 no and 0 insufficient.

CO-CHAIR BRISS: So from here, measure properties, reliability and validity?

MS. FRANKLIN: That's correct.

CO-CHAIR BRISS: So onto reliability.

DR. EINZIG: So looking at the reliability, this is a survey, and I apologize. I wasn't prepared to go over specific data on this. If anyone is prepared with specific data?

CO-CHAIR BRISS: Are these kappas that are in front of us? That's probably a developer question. The numbers that are being -- the reliability testing numbers that were being shown on the screen, I'm sorry, we needed the testing. Yeah, right. Are those kappa statistics? Are those agreement?

MS. ALAYON: No, these aren't kappa statistics. So yes. So we're using a beta binomial model. So this is different from the previous measures that we reviewed today.

CO-CHAIR PINCUS: So what is the methodology that you're using, I mean in terms of the actual method by which the data are collected and prepared?

MS. BARTON: The method by which the data are collected is a CAHPS survey, which is

administered to a sample of health plan members, depending on the potential population of the health plan.

I believe that they are phone surveys. There are experimentations currently, and so in the future, I think you'll see some variety of survey methodologies.

But for now, I believe this is a phone survey, and the reporting sequence for these smoking measures is that the patients who, members who report tobacco use are then asked "were you advised to quit," etcetera, the next questions after that.

CO-CHAIR PINCUS: The question was how did you conduct the reliability study?

MS. BARTON: The reliability study, the methodology that's used by NCQA is a statistical approach to looking at the spread of performance. So given, for example, a first year's -- a single year's administration of the CAHPS survey, taking all of the survey responses, and looking at how responses are

grouped within plan and between plans, to determine the degree of, I guess, the fineness of the knife, as it were, the degree of distinction between the plans that are being compared to each other, because this is ultimately an accountability measure, to assess the capacity of a plan in all the variety of ways it may extend a message to its members, as to how far its reach extended, and then to compare plans to each other in these public reporting venues that Dawn has mentioned.

DR. BURSTIN: Can I just help with one quick thing? So NQF allows reliability testing at either the data element level, which is what we've been talking about most of the time this morning, or for large data sets, testing at the measure score level.

In that instance, that's what they're doing here, correct me if I'm wrong NCQA, which is that they're actually looking at the signal to noise ratio of the data set and the results; is that correct, Mary and Dawn?

CO-CHAIR BRISS: So does that mean that --

(Off record comments.)

CO-CHAIR BRISS: Or comparing plan to plan variability to within-plan variability? Is it the latter?

MS. BARTON: Yes, it's the latter. It's not year to year; it's plan, within, between plans.

CO-CHAIR BRISS: Yeah. I'm not sure that I fully understand it either. It looks like some methodology where they're comparing plan to plan variability and correcting for within-plan variability or something like that.

MS. BARTON: No. Actually, there's no correction for within-plan variability. It's merely a measure of the capacity of this metric to distinguish meaningfully between plans.

CO-CHAIR BRISS: So they give, in the paragraph above Testing Results there, they give the signal to noise ratio, and they give their

perspective of what the numbers mean. So zero is bad and .7 is considered very good, and the numbers themselves are closer to .7.

CO-CHAIR PINCUS: Let me ask that question (off mic).

(Off record comments.)

CO-CHAIR PINCUS: The range of response rates.

MS. BARTON: We're looking that up now.

DR. BURSTIN: The response rate would be the same as the health plan CAHPS. It's basically questions incorporated into health plan CAHPS; correct? Yes. That's not a separate survey.

(Off record comments.)

CO-CHAIR BRISS: Vanita, were you trying to get in on this point?

DR. PINDOLIA: I had a question. So I know from CAHPS and looking at it from five-star ratings and health plans get rated on that. I'm not familiar, particularly with

NCQA, if you had other HEDIS measures pulled from CAHPS surveys.

But my experience with them, with the CMS, there's just such a variability because it's always given the first quarter. It's a mailed survey, and if the patient just got discussed with the doctor three months prior, they remember.

But when it happened nine months earlier, so like right now they have a CAHPS score for exercise, where it was "Was exercise discussed with you?" It varies so much year to year, based on when that discussion occurred. But then when we look at the actual charts for the staff model physicians where we have access, it's clearly been discussed, but the patients just don't remember nine months later, when they're asked in a survey.

Do you know? Has this been looked at NCQA? I'm not aware of them using CAHPS for others.

MS. BARTON: There are several

measures that are used in the STARS rating that come from the CAHPS survey of plans, and this is one of them. I think the precise issue that you raise is the reason why there's currently experimentation afoot to alter the methodology of administering CAHPS, and in fact, as you mentioned, it's a mailed survey.

There has been a request from the provider community to have, instead of once a year, to have more frequent administrations, like quarterly waves of surveys, so that you could catch people within their memory, hopefully plus or minus within their, you know, recent memory, to be able to recall what had happened to them.

So that, I think you can see that that's where the field is moving very quickly.

DR. PINDOLIA: So is that where this is going to go? When this gets endorsed, the way you're submitting it, it looks like it's still the current CAHPS process; correct?

MS. BARTON: Any change in

administration undergoes pilot testing, and so what we're currently proposing is what we know to be the current format.

The fact that there's experimentation being undergone, I think, is an indication of the interest in moving in that direction.

But we're not ready to specify a T's crossed and I's dotted version of that administration methodology until we've tested it.

CO-CHAIR BRISS: Yes, Bonnie.

DR. ZIMA: Okay. So just to clarify, right now, it's sort of a work in progress, aligning the time periods of these two variables?

MS. BARTON: The CAHPS survey has a time period that's in the survey. So it asks did you, in the last, and I believe that the precise item is -- we'll get the precise item in just a second. But it's standard within the survey. So it doesn't change. It's not a work in progress.

The fact that there will be waves of surveys administered in the future potentially, it is my understanding that still, that will not change the wording of the survey, but just that people will be surveyed closer in time hopefully to the visits that they got them on the list to be surveyed.

DR. PINDOLIA: So I understand. So it's not the survey, but it's the analysis, that right now the time frames are not -- it's variable, the alignment, right, between the CAHPS survey and these other HEDIS measures.

MS. BARTON: I'm afraid I don't understand.

DR. PINDOLIA: I think I was simply following up on an earlier point, that the time may not align, so that the risk of recall bias varies.

DR. CARNEY-DOEBBELING: The CAHPS is done at the same time year over year, and HEDIS is collected over the first half of the year, with a deadline of June 1st. Correct me if I'm

wrong, but the CAHPS survey has to be completed by April 1st of every year. It goes into the field. The phone calls go out typically in the month of March.

So there is recall bias that can't be ignored, but it's likely that that recall bias is the same for the members of the Plan A as it is for Plan B. So it's all a wash at the end. So yeah, there is recall bias, but it's the same for everybody who's being -- right.

CO-CHAIR BRISS: And right now, we're talking about reliability testing, right, and so --

DR. CARNEY-DOEBBELING: I just -- I know the question is a yes or a no. I think it's a yes/no and then I don't know, if I recall the wording of the question. I can't find it on here, but you were asking that earlier.

(Off record comments.)

DR. CARNEY-DOEBBELING: Oh, the response rates. When our health plan does a CAHPS survey, it's all telephone. We're lucky

to get 35 percent. That would be great.

CO-CHAIR PINCUS: Is there a difference in response rates to these items, as compared to other items on the CAHPS?

CO-CHAIR BRISS: So do we have enough information to close on reliability? So let's try --

DR. CARNEY-DOEBBELING: I think that maybe the bigger reliability question for some of the folks in the room is has it ever been tested? If I am contacted in January and then I'm contacted again in March, will I give the same answer ostensibly?

That kind of reliability, I think with these questions, has not been tested or at least is not reported here, in lieu of just looking to see how the data aggregate across all of the plans for the signal to noise.

MS. BARTON: That's an excellent point. So the CAHPS survey has been, every element on every CAHPS survey has been tested for the reliability. Every item, put it that way.

So AHRQ coordinates the CAHPS surveys that go to clinician groups, that go to hospices, that go to hospitals, and the degree of psychometric testing and research that's done on those items, before they are allowed to be incorporated into the CAHPS survey, is not included in this, but is robust.

The measure that's created from the CAHPS items is what we've reported on, in terms of the reliability, using the metric that's described here, in order to determine whether it's worth the squeeze of comparing one plan to another, using this measure that we've created from the items.

CO-CHAIR BRISS: So let's try to vote and see what happens.

MR. WILLIAMSON: We will now vote on reliability. You may begin voting now. Okay. So we have 1 high, 13 moderate, 2 low and 2 insufficient.

CO-CHAIR BRISS: So moving on to validity.

DR. EINZIG: Okay. Moving on to validity, it's very long. So it runs through the steps on how they determine validity. So according to the results from the folks who wrote the measure, they propose that they feel the validity is, that the survey is deemed valid.

Should we leave it at that? Ten in favor, one opposed, one abstained.

CO-CHAIR BRISS: So it seems to be -- the bottom line seems to be that there were at least two groups of experts that assessed the face validity of the measure, and they generally voted to support the face validity. Is that essentially it? So anybody have questions or concerns? Let's try to vote.

MR. WILLIAMSON: We will now vote on the validity. You may begin voting now. For validity, we have 3 high, 14 moderate, 1 low and 1 insufficient evidence.

CO-CHAIR BRISS: So on to usability.

DR. EINZIG: Okay. So moving on to usability, again this is a survey measure. It

sounds like it's -- according to the folks who wrote the measure, it appears straightforward again, that it is deemed usable.

CO-CHAIR BRISS: So any comments on usability before we vote?

(No response.)

CO-CHAIR BRISS: Hearing none, let's vote.

MR. WILLIAMSON: We will now vote on usability. Begin voting now. For usability, we have 6 high, 11 moderate, 1 low and 1 insufficient.

CO-CHAIR BRISS: And feasibility.

DR. EINZIG: No further comments on that, for the sake of time. Let's go for it.

CO-CHAIR BRISS: Anybody want to comment on that, from around the table?

(No response.)

CO-CHAIR BRISS: Let's vote.

MR. WILLIAMSON: Begin voting now. For feasibility, we have 8 high, 9 moderate, 1 low and 1 insufficient.

CO-CHAIR BRISS: So anybody want to make final comments before we vote on the overall measure?

(No response.)

CO-CHAIR BRISS: Hearing none, yes or no. Oh, I'm sorry.

DR. PINDOLIA: Again, just going back to what this will imply for the health plans, when they have this as one of their HEDIS measures to be measured up against, the problem we have with CAHPS surveys, whether it's the Medicare version or others, when it's a question of just did something happen, but not really knowing if it truly was discussed or not, there's really no way for us to make an impact to change the care for the next year because we don't know if there's a targeted physician population we should talk to, or if there's any targeted pay for performance that we can implement, because per chart review it's been discussed. So just to keep that in mind.

CO-CHAIR BRISS: So anybody else?

Questions, comments, concerns before we do an overall vote?

(No response.)

CO-CHAIR BRISS: Hearing none, let's try the vote.

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement. Begin voting now. And the measure passes, 17 to 2.

CO-CHAIR BRISS: So a couple of things. Would anybody on the phone like to comment?

MS. FRANKLIN: Operator.

OPERATOR: And that is *1 if you'd like to comment over the phone.

(No response.)

OPERATOR: No one has signaled.

NQF Member/Public Comment

CO-CHAIR BRISS: Okay, and at this point, we'd like to pause for a second and take a deep breath, and ask if anyone, either on the phone or in the room, would like to make a public comment. So maybe on the phone first, or could

we have any public --

OPERATOR: Again, that would be *1.

(No response.)

CO-CHAIR BRISS: And hearing none, is there anybody in the room that would like to make a public comment?

(No response.)

(Laughter.)

Related and Competing Measures Discussion

CO-CHAIR BRISS: Sadly for us, we don't count as the public in this context. So the last thing, our last detail to tie up on tobacco is the related and competing measures discussion.

Skip until tomorrow. So we will have that discussion tomorrow. Let's take a ten minute break before we start the alcohol measures, and reconvene at 20 til please.

(Whereupon, a short recess was taken.)

CO-CHAIR PINCUS: Why don't we get started? So I'll do this portion and you can do the next portion.

MS. FRANKLIN: Felicia, are you there?

OPERATOR: Yes, your line is open.

MS. FRANKLIN: Thanks, Felicia.

OPERATOR: You're welcome.

MS. FRANKLIN: We're looking for one of our Committee members, Dr. Naegle, who might be on the line.

OPERATOR: She had been on the line, but she has disconnected.

MS. FRANKLIN: Okay.

(Off record comments.)

CO-CHAIR PINCUS: So we're going to -- should we skip over that one, and go to the Joint Commission one? Go to 1661. Okay. So we're going to wait until we can reach Madeline, and maybe we can go to 1661. Who's the lead for that?

MS. FRANKLIN: So that's Jeffrey Susman, but we do have to hear a little bit from the measure developer, to tee this up.

Measure 1661

MS. LAWLER: This is the one of four measures in a set of measures that address screening for alcohol use, brief intervention, for those that screen positive, and treatment at discharge with referrals or prescription for medication, and then the follow-up measure.

So this is the first measure, which is the screening, and in the denominator, we're screening all patients 18 years of age and older, regardless of diagnosis. So it's not diagnosis-specific.

This a global type of measure, and in the numerator, we're simply looking to see the number of patients that were screened for alcohol use, using a validated screening tool.

DR. SUSMAN: Okay. Well, this is deja vu all over again. You'll note a lot of similarities, perhaps, to our discussion around smoking, and hopefully, for the poor folks from JCAHO, we'll be able to get a few of these passed here.

So without further ado, this is a

hospitalized patients, 18 years and older, who are screened during their stay, using a validated screening questionnaire and these are specified. It is part of a family of four measures, similar to what was discussed during our discussion around tobacco.

The first issue is impact, and I think there is a very nice summary of the high impact of alcohol substance abuse, and cost to society. I hope that we probably don't need to dwell a lot of time on the fact that individuals have a high burden of morbidity related to alcohol use.

CO-CHAIR PINCUS: Are there any comments on that?

DR. SUSMAN: And the group, the work group, by the way, which included David, Madeline, Tami, Jeff and Mady, all thought this was a high impact condition.

CO-CHAIR PINCUS: Any other comments with regard to impact?

(No response.)

CO-CHAIR PINCUS: I guess we're ready

to vote.

MR. WILLIAMSON: We will now vote on the impact. Begin voting now. For impact, we have 18 high, 1 moderate, 0 low and 0 insufficient.

CO-CHAIR PINCUS: Gaps.

DR. SUSMAN: Okay. The next is the opportunity for improvement. The work group felt in general that there was a demonstration of performance gap, with 4 high and 2 moderate or medium. The issues here is perhaps less of the data that might have been provided.

So while there's clearly a performance gap, the evidence process orientation to actual outcomes is less clear. But in any case, if you just look at the screening and the opportunity to screen, there seems to be a fairly large gap in ideal performance and current performance.

Some of the information's generalized from other locales, but I think it would be fair to say, at least in the work group's assessment, that there is an opportunity for improvement

here.

MS. FRANKLIN: Discussion on the gap?

(No response.)

MR. WILLIAMSON: We will now vote on the performance gap. Begin voting now. For the performance gap, we have 12 high, 7 moderate, 0 low and 0 insufficient.

DR. SUSMAN: And now we'll move on to the evidence issue. The issue here really is linking the screening to an ultimate outcome. Obviously, to get an ultimate outcome, you need to know what your baseline is. So as the rationale for our tobacco measures, this really sets up for measurement of depression care and improving depression.

It's a similar thing. If you don't identify, if you don't screen at the outset, it's hard to know whether you're going to have any impact. So this sets up, if you will, the group of patients who are eligible for intervention and follow-through to an outcome.

The negative side, just to be fair,

would be that a lot of the data have been generated at the outpatient setting. My own assessment of reading this in the literature is that there really is sufficient evidence in the inpatient side of the house, that this is important.

There is a clear link to evidence of an outcome. I don't think that's an overstatement, although certainly less evidence than perhaps others.

CO-CHAIR PINCUS: Comments, questions. Okay, Caroline.

DR. CARNEY-DOEBBELING: A quick question. Why was this limited to 18 year olds, instead of going younger, especially with unhealthy and/or binge drinking?

DR. SUSMAN: I think I'll leave that to the measure developers to answer.

MS. LAWLER: This was obviously a discussion that we had with our technical advisory panel, and we decided, knowing that people begin to drink at a younger age, that we

needed to stick with where the evidence was, which was largely with the adult population. So that's primary why we kept the age at 18. Dr. Fiori.

DR. FIORI: There is good evidence, using standardized screening tools, including the audit but also the craft and other instruments, that there is good reliability, validity, sensitivity, specificity for 13 to 18 year-olds.

Also, I think the main reason why we did not go to the 13 to 18 year-olds was the decision from the U.S. Preventive Services Task Force, which said that there is strong evidence with randomized control trials for 18 and above, that there is no reason not to think that it would be effective for adolescents.

So they extended their recommendation to adolescents as well as adults, but there did not say that there was the RCTs for that.

Similarly, I think there was a sentiment that other substances of use should

also be screened, and for example, the SBIRT program that CSAT runs screens for tobacco, alcohol, illicit drugs and prescription medication misuse.

But again, the Preventive Services Task Force said that there were insufficient randomized control trials. So we stayed with the evidence.

DR. SUSMAN: My own sense is by having a narrower population, you're sticking more closely where the best evidence that exists.

DR. MARK: Yeah. I just wanted to provide input on the evidence for the panel. A lot of it comes from a Cochrane Collaboration review. McQueen is the author. It's 14 studies that looked at brief interventions in general hospitals, seven of which were randomized clinical trials that overall concluded that it was effective.

CO-CHAIR PINCUS: The "it" being screening?

DR. MARK: It's actually screening

and brief intervention. So, but yes.

CO-CHAIR PINCUS: So any other comments? Oh, Jeff.

DR. SAMET: Yeah. So the issue that we raised with tobacco, I mean this McQueen, "Brief Interventions for Heavy Alcohol Users Admitted to General Hospitals" were six studies. Participants were not randomized to control or brief interventions. It's unlikely that allocation to the point of assignment was concealed.

It's sort of the stating -- maybe I could be better informed about the McQueen article, because in this one, it wasn't mentioned in this one, though it's mentioned in the next protocol we'll get to, the Saitz article, which I will admit I'm senior author on, but looked at and did not find -- yeah. I mean it's a good paper.

But the other issue here, in talking about when we look at the inpatient setting, unlike the outpatient, where with smoking I made

the point with the population, there's no inherent reason to think it's different.

It wasn't clear in this protocol either -- when you screen in the outpatient setting, you pick up 80 percent people who are at-risk drinkers, and 20 percent who might be dependent kind of roughly.

When you screen in the inpatient setting, you pick up 80 percent who are dependent, and 20 percent who are at risk. So the inherent -- so there's major sort of differences when one looks at what might happen down the line.

Now I don't know if we should talk about it here or talk about it with the next protocol, but it's going to play out with each of those. So maybe I'll stop there and just maybe get better informed.

(Off record comments.)

DR. SAMET: Well, the logical conclusion is there's a -- if you thought smoking was worrisome, this is a lot more worrisome, to

base the recommendations on outpatient data. Now we're not only doing that. You mentioned the McQueen paper.

But the McQueen, you know, kind of meta-analysis or at least review paper. But that's why maybe if there's some more data on McQueen that you can relate to, that would be helpful, because there's at least a study which was a very nicely-controlled randomized control trial which unfortunately, you know, or fortunately, whatever, didn't show the difference that was being looked for.

Part of it might be explained by the fact that you're trying to do a brief intervention on people with alcohol dependence, for the most part, and that's not so doable. And even linking them to care, which would be a great benefit, didn't reveal surprisingly that finding.

DR. FIORI: There are two McQueen Cochrane reviews. The 2009 review was the one which was quoted, I think, in the next study, as

staying that there was not strong evidence or not incontrovertible evidence. Or let me quote the 2011 Cochrane review:

"Fourteen studies involving 4,041 mainly male participants were included. The results demonstrate that patients receiving brief intervention have greater reduction in alcohol consumption, compared to control groups at six months and at nine months follow-up, but this was not maintained at one year.

"Self reports of reduction of alcohol consumption at one year were found in favor of reductions, in addition to significantly fewer deaths in the group receiving brief interventions than in the control group at six months.

"Furthermore, screening, asking participants about their drinking patterns may also have positive impacts on alcohol consumption levels and changing in drinking behaviors." So what they found was 12 out of 14 studies.

"Also, a report by Nilsen, a systematic review of emergency care brief interventions for injury patients, the results of that, of 12 studies that compared pre-post BI results, 11 observed significant differences of BI on at least some outcomes, alcohol intake, risky drinking patterns, alcohol-related negative consequences and injury frequency. More intensive interventions yielded more favorable results."

So the 2011, and then also this 2008 Nilsen et al. systematic reviews find that not all studies find positive outcomes, and also importantly, not all studies that find positive outcomes find that it's effective for everyone.

We also have some evidence from work by Craig Fields and others, that the screening and brief intervention is effective across ethnicities, at least for Caucasian, Hispanic and African-American.

CO-CHAIR BRISS: And those were -- all of those studies were inpatient studies?

DR. FIORI: They all are general hospital inpatient studies.

CO-CHAIR PINCUS: Just a clarification. Did some (off mic).

DR. FIORI: Well, it's emergency patients who go through the ED into trauma care.

(Off record comments.)

DR. FIORI: Yeah. So they end up as -- now another issue which -- and this is not in any way picking apart the Samet-Saitz study, I particularly wouldn't pick on that when somebody voting is the author, but there are likely very different prevalences of both risky use and dependent use, depending on the service.

So trauma care is likely to have a much higher rate. The psych unit is likely to have a very high rate of dependence. An OB/GYN unit would have low. A GI unit would have high. Likely there's considerable variability in the type of unit and the type of hospital.

(Off record comments.)

CO-CHAIR PINCUS: Was there a segment

in the analysis done of just the people with abuse and not dependence, to look at outcomes?

DR. SUSMAN: Yeah. There was analysis done of the risky drinkers, but 23 percent of the study had only at-risk drinkers. So you know, it wasn't powered to look at such a small group, and it didn't show anything. But I can't really say much about that.

Yeah, and you know, it kind of post hoc showed that there were some groups that actually did benefit, you know, if you pick out an age in the thing. So you know, I would have loved for the results to have been otherwise, to be honest with you, but they were what they were.

(Off record comments.)

MR. WILLIAMSON: Yes, we will now vote on the evidence. You can begin voting now. We have 17 yes, 0 no and 2 insufficient.

DR. SUSMAN: Okay. So moving right along. This is now on the reliability and validity issues. First of all, I think everybody noted that there were some reliability

issues. In the end, our work group was split between moderate and high, mostly moderate, with 5 members voting moderate and 1 high.

If you look at it, the reliability in the validation sample was .252, which you know, it depends on if you're a cup half full or half empty type of individual. There certainly the issue of abstraction and saying that someone was actually screened on an unvalidated screener, in other words, using something that wasn't one of the vetted tools for doing a highly robust screening.

So that was the big issue, as I understand in reading the materials. I don't know if the measure developers want to comment further on that.

MS. WATT: No. You're correct in that assumption.

DR. SUSMAN: So like much of our discussion today, you know, yes, the reliability could be improved. I don't know if there were any changes made to the measure with that

particular issue surfacing.

MS. WATT: Yeah. I just to address a different point, not your question, sorry.

DR. SUSMAN: Yes, Seneca.

MS. WATT: Based on what I recall, we did ask our statisticians to run, to look at these, and what they did is they included a 95 percent confidence interval.

So for this first one, although it's not great, if you -- the confidence intervals runs out to .423. This is a small number of cases, and so there's a very wide confidence interval.

I don't know if that helps or hurts, but I wanted you to be aware of it, because that's actually true for all of these. Nancy, did you want to address some of the specification strengthenings that we did?

MS. LAWLER: No. We felt that -- we felt primarily that it was just the issue of getting used to using a validated tool, as opposed to something that the hospital made up

themselves. Did you have something to add?

DR. FIORI: Yes. A couple of things. We required the use of a standardized instrument. The reason for that is that when you compare standardized instruments with either the research diagnosis or DSM diagnosis, you get sensitivities and specificities of .8 plus, with the standardized instruments.

If you rely on ad hoc, you get a couple of really very bad things. First is you get very poor sensitivity, and you really get to see what the biases are of the providers. We tend to get very high rates of false positives among young African-American, disheveled males.

We get very high rates of false negatives of white, middle-class, older, well-dressed females, and it's very systematic. That gets picked up by very brief standardized instruments.

DR. MARK: I think we discussed this, but I can't remember exactly what the outcome was. It seemed like some hospitals were using

a two-stage screening, where they'd ask did you drink at all, and then if they said yes, then they'd go to the second stage.

What's your justification for not allowing that kind of two-stage screen to count towards screening, because it sounds like you would count that as not screening? So you'd have to use an AUDIT for everybody. You couldn't use that two-stage approach.

DR. FIORI: The second screen counts. So asking "do you drink" is not screening. But if then you ask sort of three questions of the AUDIT-C, that counts.

The two-stage does count. It's the asking of "you don't drink too much do you?" doesn't count.

(Laughter.)

DR. NAEGLE: Yeah. Hi. It's --

MS. FRANKLIN: Who is this?

DR. NAEGLE: Madeline Naegle. I'm on the line. I just wanted to -- I unfortunately had to be out of the meeting for a period of time.

We have been using a two-stage process in our collegiate population, and we have found that to be very successful, to ask them how many times they had had more than the recommended NIAAA level say, or how many times they had had more than three or four drinks in the recent month.

Then if they were positive on that, we went ahead to using the AUDIT-C. That just speaks to the individual who made the comment about the notion of two-stage, when can two-stage be effective. We have found that to be very helpful in our depression, our collegiate depression screening project, to which we added screening for alcohol misuse and abuse.

CO-CHAIR PINCUS: So Tami, do you want to comment on that?

DR. MARK: So I guess I'm just not clear when you did the reliability test, if you -- how you measured, how you captured people who only did the first stage and said they don't

drink, and then went on to the second stage. Does that count as screening?

MS. LAWLER: In the reliability studies, in the first, the specifications as we tested them, we didn't account for anything other than using a validated tool. When we got out there, we realized that a lot of people were doing this two-stage process.

So we actually made some calls back to Eric, to find out, you know, is this, you know, can we allow this? Is this appropriate, because it's not really the way that we set up the measure. We did, in our final specifications then, allow for this two-step methodology to occur.

CO-CHAIR PINCUS: So could you maybe explain a little bit more what this -- is the two-step methodology what Tami described or what Madeline described? It seems to me -- or what Eric described?

DR. SUSMAN: What is a two-stage screening? What qualifies for it in --

MS. LAWLER: I think if they use the single validated question from the NIAAA, oh I'm sorry, the single validated question from the NIAAA about do you -- I can't remember exactly what it is. Eric, you probably know it by heart.

DR. FIORI: It's the binge drinking question, five drinks or more.

(Simultaneous speaking.)

MS. LAWLER: And then we find out that the person is using alcohol. Then they can go on to the validated tool, to do a further assessment.

DR. KHATRI: So we do this pre-screen, and we call it a pre-screen. So we use that one question. Then we have the pre-screening, and if they say yes to that, then we move to the validated measure. But if they say no to that, we stop.

So the question is if someone says no to that, does that count as a screen, even though you use the pre-screen? We did not go further. So in terms of workflow, it just really is much

more efficient to do the pre-screen. But you should account for that in your numbers.

(Off record comments.)

CO-CHAIR PINCUS: So the question is do you allow for a formal pre-screen, i.e. a two-step procedure, and if somebody says no to the first one, does that count as being positive? And number two, if that is the case, was that the way in which you dealt with it in compiling the reliability data?

MS. LAWLER: Okay. It would count as, because it's a single validated question. So it counts as a validated tool. So we would count that. Did we use that in the reliability testing? No, we didn't, because it wasn't part of our specifications at that time. It is now, but it wasn't at that time.

DR. CARNEY-DOEBBELING: So the specifications have been redone but not retested; is that correct?

MS. LAWLER: We revised the specifications based on the findings of the

reliability study, and no, there was not -- it as not retested.

CO-CHAIR PINCUS: Lynn.

DR. WEGNER: Going back to the issue of screening, there's a model for this in developmental screening and surveillance. Surveillance is asking a question or just having concerns.

Screening is very specifically defined as the administration of a standardized instrument, and it's not testing, which is a formal assessment, but it's a validated instrument, and actually it's part of CPT-96110, which is developmental screening. It specifies the use of a standardized, validated instrument.

So I think that would answer the question. A pre-screening question would just be surveillance.

CO-CHAIR PINCUS: Are there other comments, questions on reliability?

DR. SUSMAN: The one thing I'd just add again is it would be very helpful to do

reliability testing, once the measure is locked. But I know you're sort of in a no-win situation, to some extent. But getting the results on a measure that's been changed in the specification doesn't give me a warm, fuzzy feeling.

I think the other thing that I keep seeing, is relatively small validation sample, 131 individuals. I guess -- well, I know it costs more and there's a lot of issues around that. I just wonder if a lot of the questions we have around this table would be mitigated, if you didn't do it on a larger sample.

CO-CHAIR PINCUS: Any other comments with regard to reliability?

(No response.)

CO-CHAIR PINCUS: Okay. Let's vote.

MR. WILLIAMSON: We will now vote on the reliability. Begin voting now. We have 1 high, 8 moderate, 4 low and 6 insufficient.

CO-CHAIR PINCUS: Do we stop?

MR. WILLIAMSON: The measure fails on scientific acceptability.

CO-CHAIR PINCUS: Okay. So that stops this measure. Should we go back to the one -- Madeline? Madeline are you there?

DR. NAEGLE: Hi there, how are you? Yeah, I'm on.

Measure 0004

CO-CHAIR PINCUS: So we're going to go back to the measure, Angela, what was the number?

MS. FRANKLIN: 0004.

CO-CHAIR PINCUS: Okay.

MS. FRANKLIN: That one was, the title is Initiation and Engagement of Alcohol and Other Drug Dependence Treatment, and we'll have the developer tee it up for us.

DR. NAEGLE: Yes. Is Michael on today?

MS. FRANKLIN: Michael is not on today. He wasn't able to make it.

DR. NAEGLE: Okay, all right.

CO-CHAIR PINCUS: Jeremy, are you going to be --

MR. GOTTLICH: Yeah. I'll introduce

the measure. So we're looking at the initiation and engagement of alcohol and other drug dependence treatment. This is a measure that is looking at the percentage of adolescent and adult members, with a new episode of alcohol or other drug dependence, who received one or two different things.

The first is an initiation visit, which is calculated as a visit within 14 days of a new diagnosis. The second rate is an engagement visit, which is calculated as two visits within 30 days after the initiation visit. So as they get into the engagement rate, you need to have first fit in the initiation rate.

So there's two rates for this measure. This is a health plan measure that's specified for commercial, Medicaid and Medicare plans. We have two age stratifications within the measure, 13 to 17, which is the adolescent rate, and an 18 and older, which is the adult rate.

We also calculate a total rate for the

measure for both rates. We require two benefits that the health plan must have to report this measure. That's a medical benefit and a chemical dependency benefit. The chemical dependency benefit must be inpatient and outpatient.

Just to explain the benefit a little bit, an organization is -- a health plan is responsible for reporting at HEDIS measures, which they offer the benefit directly. Organizations are not responsible for reporting members that do not have the benefit.

This has been included in the HEDIS measurement set since 2004, and since 2004, it's been reevaluated by our Behavioral Health Measurement Advisory Panel several times, as well as being reviewed by our Committee on Performance Measurement.

Some of the things that they've changed during the time involve coding, also combining elements as initiation and engagement, and then also I believe the 13 to 17

age range was added after the measure was first introduced.

CO-CHAIR PINCUS: Okay. Madeline, do you want to walk us through the process, starting with importance?

DR. NAEGLE: Yeah. Certainly, this is an important measure, inasmuch as it relates to access, which often doesn't exist for individuals, even when they are covered by health plans. So the goal is to increase access and quality of care.

The impact in our work group, we discussed the importance as being -- I think you have those ratings. Certainly, I felt that it was a high importance in the potential to affect two population groups, adolescents and adults.

Going on to -- so that's impact. Going on to the evidence. Harold, do you want to discuss impact --

CO-CHAIR PINCUS: Wait. We need to make -- first, we need to have a vote on importance.

DR. NAEGLE: Uh-huh, importance and impact.

CO-CHAIR PINCUS: Yeah, importance and impact.

DR. NAEGLE: So the data that is cited about interventions and successful outcomes for interventions essentially comes from guidelines, and information that we know about people's potential once they're intervened with, to have good recovery outcomes. But at this particular point in time, we did not have studies looking at any of the groups who were identified immediately post-intervention on this particular measure.

CO-CHAIR PINCUS: So are there comments with regard to impact? Questions, comments, with regard to impact?

(No response.)

CO-CHAIR PINCUS: Okay.

MR. WILLIAMSON: We will now vote on the impact. Begin voting now. For impact, we have 15 high, 4 moderate, 0 low and 0

insufficient.

CO-CHAIR PINCUS: Okay. Can we talk about now opportunity for improvement Madeline?

DR. NAEGLE: Uh-huh. The measure looks at the degree to which the organization initiates and engages members identified with a need for alcohol and other drug treatment. The possibility for improvement, in terms of the database and what we have been looking at from a scientific perspective, suggests that the evidence is fair to moderate, that individuals will be able to respond effectively.

But there, in looking at the data that has been proffered with the outcomes and implementation of this so far, that is not as strong. So the recommendation was for moderate, from my perspective, and actually opportunity for improvement is high on this. My recommendation from the work group.

CO-CHAIR PINCUS: Any comments, questions, with regard to --

DR. NAEGLE: Harold, I can't hear you.

I know you're off mic, but --

CO-CHAIR PINCUS: Sorry, sorry. Opportunity for improvement. Any comments or questions? We have a comment or question from David Pating.

DR. PATING: I'm going to put it out there, but I have problems with this measure, primarily because the definitions that it uses are very slippery. Ostensibly, it's measuring whether patients with dependence make it into treatment, they get initiation and engagement into treatment.

But when they start to demonstrate the gap, they start talking about patients with abuse, patients with dependence, patients with substance use.

It's very unclear from the evidence they provided that they're really talking about a homogeneous population of connecting people that need treatment into treatment, based on the studies that they used.

Then with later, I think it's 1(b),

looking at the summary of demonstrating performance gap, they use their own data, which is sort tautologous, basically that our data shows that there's a gap. But I'm going to tell you quite honestly, when you get into the definitions of what is in the numerator of the measure, it's not quite obvious that what you're measuring is what you're getting.

The measure measures both abuse and dependence. It's called dependence initiation and treatment, but there's a lot of other stuff, both in the literature evidence that they provided, as well as the numerator that will be supplied later on, that makes this not a true measure of dependence and addiction needing treatment.

DR. NAEGLE: So you're talking about the disconnect -- I'm sorry. The disconnect between what's identified in the numerator in very general terms about admission and detoxifications, and then the making of appointments as supported by records and

documentation?

DR. PATING: Well, I'm actually talking on two levels. One is on this performance gap issue that we're looking at. They cite several studies. One of them they say SAMHSA 2011, 20.5 million persons classified needing substance use treatment did not receive treatment.

One million felt they needed treatment for illicit drug or alcohol use problems. Is that problems like abuse, or is that problems like dependence?

DR. NAEGLE: Okay.

DR. PATING: Then later on, they say again less than 20 percent of those with substance abuse needed treatment, and less than 40 percent of addiction needing treatment.

There's these conflicting -- there's this conflation of terms throughout the definition of this measure, that really makes the process of what are they capturing with this, you know. It undermines the validity.

So I'm just saying with regards to the gap, the gap that you really don't know what you're getting when they're reciting this gap literature, because they're citing their own measure as a measure that there's a gap, that they're showing that -- we're showing 40 percent of people are getting screened.

CO-CHAIR PINCUS: So let's see if the measure developer can clarify how the numerator is being defined, in relationship to both the title of the measure, as well as how the data's put together around the gap?

MR. GOTTLICH: Well, I'll try to give it a go. So we have an initiation engagement rate. So we're really looking at the coordination of care in this measure. Are members getting access to the care that can be available within an immediate period after their first diagnosis of alcohol and other drug dependence?

As far as the type of coding we include that can get you into the measure, this includes

dependence codes and it includes abuse codes and also, I think there was mentioning about the 305 code, which is non-dependence treatment.

So I just want to clarify that this measure really is an access paired process measure. We see it as an intermediate measure between -- that comes before a settled outcome.

DR. NAEGLE: So it's certainly not seen as a matching process? So it's really an indicator of people being noted at all in the system, with a range of codes that you've given us. Is that correct?

MR. GOTTLICH: That's correct, and then we have two levels, both the first initiation within 14 days, and then 30 days after that, if they have received more intensive follow-up, which is considered two or more visits.

CO-CHAIR PINCUS: So let me just clarify then, what I think is happening. I think David, if there was some greater consistency with regard to their description of

the database around importance, and that it was clearly parallel to how they're describing the gaps, which in turn was clearly parallel with how they're defining the numerator and denominator, and in turn with the title of the measure, you would be more comfortable?

So is it a matter of sort of the consistency of the evidence attribution and the actual definitions and what the measure's called?

DR. PATING: Well, that's where the problem is first showing up, because we're taking this sequentially. I'll tell you the bigger issue is really with the reliability and validity of the measure.

(Simultaneous speaking.)

DR. PATING: But right now, the performance gap I would rate probably more in the moderate, low-moderate range, based on the evidence that they're providing.

But there's sort of an internal tautologous argument going on here, that we just

really want to make sure these slippery definitions that we're aware of. I'll bring it up under the reliability, the bigger issue, in a minute.

CO-CHAIR PINCUS: But I think there is the issue of the title of the measure says "dependence," when in fact it's not just dependence.

MR. GOTTLICH: I think we might get to it later with reliability, with the coding. But we are, when we look at transferring the ICD-9 coding to ICD-10, that is an issue we are looking at, as far as how it transfers to a use, abuse and dependence.

That's going to be where we focus on really refining the measure, making it more of a dependence measure than just any abuse code.

CO-CHAIR PINCUS: So any other comments with regard to (off mic). Privately. (off mic)

MR. WILLIAMSON: We will now vote on the performance gap. Begin voting now. All

right. We have 5 high, 10 moderate, 3 low and 1 insufficient.

CO-CHAIR PINCUS: Madeline, let's move on to evidence.

DR. NAEGLE: Okay. So looking at the evidence, beginning with reliability, again there is some confusion about terms with the support that has been cited. I introduce, since the denominator statement there is also the use of episodes, as opposed to actual coded outpatient/inpatient --

CO-CHAIR PINCUS: Actually Madeline?

DR. NAEGLE: Yes.

CO-CHAIR PINCUS: Actually, if you could go to sort of 1(b)(5) and 1(c) is where the part is. It's before we get to reliability.

DR. NAEGLE: Okay, sorry. 1(b)(5).

CO-CHAIR PINCUS: And 1(c), 1(c)(1).

DR. NAEGLE: Okay. Based on the NQF descriptions for rating the evidence, the developer's assessment. No, that 1(c). Okay.

CO-CHAIR PINCUS: 1(c)(1), structure

process outcome relationships.

DR. NAEGLE: Okay. I'm going to hand that back to the Committee. I don't seem to have it before me.

CO-CHAIR PINCUS: Okay. So I don't know. If Mady or -- if you want to comment on the evidence base. So I mean basically, is the -- basically, I mean, from my knowledge of this, basically is a finding that the longer people stay in treatment, get engaged and stay in treatment, the more likely they are to be successfully treated.

DR. NAEGLE: But yes. I did not find that that particular argument, however, was strong in relation to the identification of people's ranking and coding, and the fact that they would both initiate and stay in treatment, where the evidence was strong that there was a relationship between involvement in treatment and outcomes. So I did not find that the evidence was strong in that regard.

DR. CHALK: I don't think this is

viewed as an outcome measure yet. I think this is a -- this is Mady -- this is a process measure that's looking at do people get identified, and do they initiate and get engaged in treatment.

That really is a care coordination, in addition to whether they're identified, screen. Do they actually move from there to being engaged in treatment? That's all this is about, right?

CO-CHAIR PINCUS: It's a narrow issue. It's a fairly narrow process measure of essentially the early frequency of treatment.

DR. NAEGLE: Okay. So looking at the summary of data under 1(b), 1(c) and some of the studies which are cited there, there seems to be an omission on data about the capacity to identify and engage people in treatment.

DR. CHALK: 1(c)(15) has a little citations of evidence other than guidelines. There are a lot of studies.

DR. NAEGLE: Yes.

DR. CHALK: A considerable number of studies that have looked at the whole issue of

initiation and engagement, and with regard to the outcomes issue, that's down the road a bit and is being worked on now. But I don't think it's relevant to this discussion.

CO-CHAIR PINCUS: But the question is what's the association of this measure with outcomes? Jeffrey, did you want to say something?

DR. SAMET: That was my point, that it is relevant to just the point you just made.

DR. PATING: Well, they do cite the Harris study, 2008, and there's actually a 2009 and a 2010 and a 2011 version. So actually it would have been nice to have all the sites there, with Harris and Humphreys out of Menlo Park.

It's interesting. While the measure is poorly constructed, those that do get engaged do result in having lower ASI scores, which is Addiction Severity Index scores down the line. For the engagement score, the measure does not pan for the initiation score.

So there's probably, on a -- what the

authors say is that on an individual clinical level, there's some basis in logic about getting people engaged that seem to do better.

What the authors and the point that I'd like to make is say is that on a population basis or a facility basis, as an accountability measure used for health care systems, weighing one system or one facility versus the other, there's such variation in this that it may not be a useful measure.

But there's some, on an individual basis, linkage of individual engagement to an outcome on an ASI. So, and that's here cited under 2008. That was *Health Services Research*. That was a good journal.

CO-CHAIR BRISS: Your rationale for that statement is because we're talking about small numbers of people? I'm trying to get my head around if there's an -- if at the individual level there's an association between this, the starting the treatment cascade and better outcomes. Then why wouldn't that also be true

at higher levels of aggregation?

DR. PATING: Well because again, it goes into the reliability issues. But then there are 2010 and 2011 studies. When they start actually looking at --

When the diagnoses are made in substance abuse clinics, there's high correlation with the CPT code, which is one end of it, and the diagnosis matching up when they do chart audits.

Yes, they did get the addiction visit 90 percent of the time. When the diagnosis seems to be made in non-substance abuse clinics, there's a high false positive rate, that they got supposedly a visit type, a diagnosis and a concordance rate of only 62 percent, dropped down from 90 percent.

What they're saying is that there's something that's happening with these CPT codes across systems, that is just not panning out, as well as --

And what I'm also telling you is that

when you start throwing abuse diagnoses in there as well as dependence, and you start measuring how systems, some are measuring more abuse than dependence, and some are using -- that are CPT code and billing-driven, you're going to get large variations across systems as an accountability measure.

But on an individual basis, somebody that had the diagnosis, got engaged, did show their ASI, seem to get improved. Make sense?

(Off record comments.)

CO-CHAIR PINCUS: Then I want to ask the measure developers just to explain sort of their understanding of this data and to respond.

DR. SUSMAN: Yeah. I mean I think we're confounding the issue of evidence and the issue of the reliability and validity.

(Off record comments.)

DR. SUSMAN: Yeah. I mean so for me, you know, the evidence of early engagement, early intensity of therapy is moderate. It's not 100 percent. It's not real, real strong,

but it's simply suggestive. It would be strong enough for me.

So in that rating, I'd say probably it's a moderate. When we go on and talk about the reliability, I have some real concerns. But I think we need to stick to the evidence discussion.

CO-CHAIR PINCUS: Other comments about the evidence, and did you want to comment on the evidence, in terms of the overall rationale, from the point of view of NCQA?

MR. GOTTLICH: I'm not sure what the question is. I can say a lot of this evidence has been -- we've received and looked at. This measure's been around for seven years, and this is the evidence we've looked at during the reevaluation of the measure.

As far as the initial evidence for the creation and development of the measure, that was done by the Washington Circle Group, funded by SAMHSA. I'm not exactly sure what their evidence was back in the development of the

measure.

CO-CHAIR PINCUS: Any other comments? So let's vote on evidence.

MR. WILLIAMSON: We will now vote on the evidence. Begin voting now. We will revote. Now we're on the evidence. Just a reminder, this is a yes, no, insufficient question, 1, 2 or 3. Please begin voting now.

We have 17 yes, 0 no and 2 insufficient.

CO-CHAIR PINCUS: Now let's move on to reliability. Madeline.

DR. NAEGLE: Yes. So looking at the reliability testing, and I had some questions about this in terms of the HEDIS measures. But it would seem that it's been in place for seven years, and the estimates, as proposed, that the reliability of the system itself would support the notion or the concept of this measurement.

I think that other people in our work group had some other questions about that. I think David, you wanted to speak to that?

DR. PATING: So the point that I'm making again is this is a very public measure. I have no qualms with the concept or the evidence, and certainly we want to get people into treatment and earlier treatment and durations of treatment do affect outcomes.

But I don't think it's -- my problem with this measure is that there is wide systemic variability across health systems, and across facilities, even within one system, as shown by the VA evidence.

And it also includes diagnoses that I think are perhaps more appropriate. Rather than referral to specialty care, which this mandates, you get initiation in specialty care, I think many of the abuse or intoxication diagnoses are perfect examples of what should be getting SBIRT kinds of initiatives and brief intervention.

So this initiative actually works against what's a national trend, moving towards more integration into primary care, and takes

those diagnoses and says no, the way to score high is to send those off to specialty care.

So again, depending on where your system falls along that domain with the diagnosis, that's one issue. But just the overall use of the CPT codes and the variability creates another level of reliability problems.

DR. NAEGLE: David let me --

CO-CHAIR PINCUS: Madeline.

DR. NAEGLE: I just wanted to ask David about looking at the minimum number of visits as two. Would you see that as eliminating the possibility of that individual being kept in a primary care setting? I mean two visits would be maybe the specialty minimum. You're really talking about the potential of bypassing a viable system for treatment to go to specialty care, are you not?

DR. PATING: Well, the index visit is anything that's like -- any alcohol with intoxication, withdrawal-related symptoms, any abuse, any dependence. This measure measures

that you're supposed to get them from wherever they are in your facility into a specialty care appointment, with a certain CPT diagnosis combination.

And what the VA study is showing is that if those diagnoses were made in non-substance abuse clinics, there was low overall validity, in terms of whether the people that got the codes -- a lot of false positive rates.

So a lot of variability on the applicability of that, as measuring system against system. That's how this measure is being used.

It's a very publicly reported measure, and it's not an apples to apples kind of thing. So I just really want to make sure that we're very explicit about the problems with this measure.

The concept is great, but again, it doesn't measure dependence. It says abuse should be going to specialty care, and with the CPT diagnosis combinations, the VA studies have

shown there's a lot of false positives.

(Off record comments.)

CO-CHAIR PINCUS: Madi, Tami and (off mic).

DR. CHALK: I don't read this measure at all that way. It says once identified, another visit. It does not specify specialty care anywhere. It says "another visit," and then it says another visit after that, two visits.

DR. SUSMAN: Well, could the measure developers respond?

DR. CHALK: Do you want to respond?

MS. ALAYON: If we're invited to respond, we'd be glad to. So I think that, David, the point that you have brought up reminds me that claims are imperfect, and that since we're at this point still working with 20th century technology in terms of the majority of measures for health plans that are based on claims, we have to deal with a tremendous lack of specificity.

If we were holding individual clinicians to a performance achievement, I think I would run your argument all the way down the road. But since we're not, since we're talking about a population measure for health plans, and we're limited to working with claims data, and there is a highly variable use of claims.

There's a highly variable use of codes in visits across the country, across counties. So I actually think that by being broader, and including abuse as well as dependence, we are creating a space for an apples-to-apples comparison, ironically.

I think that what we're doing is saying we're going to make it general enough, that if you've gotten somebody's -- if you've crossed somebody's threshold as being someone with an alcohol, potentially with an alcohol problem, we want to include you in the denominator.

We don't think that this is a measure that where 100 percent is going to be the appropriate place to fall out, because that

would probably be maybe sending some people who just had one weekend bender to treatment, and maybe that wouldn't be appropriate.

But that for the purposes of comparing one health plan's population to another health plan's population, this was the best that our experts came up with, to use claims to be able to identify a population potentially at risk, and that's where the measure development over a population, I think, differs from what you would want to hold an individual clinician responsible for.

CO-CHAIR PINCUS: Tami and then Caroline.

DR. NAEGLE: But --

DR. MARK: You just clarified the issue of -- it's not specialty treatment.

MR. GOTTLICH: So in the visits that you can have in the numerator, we allow for visits in inpatient settings, outpatient settings, and it goes along with the chemical dependency benefit. We also allow for

intensive outpatient and partial hospitalization visits. Any detox visits are excluded from the numerator.

DR. MARK: So my understanding is you just need the diagnosis code to show up twice. You don't need it associated with a particular CPT code?

MR. GOTTLICH: You need the diagnosis to show up with a visit.

DR. PATING: No, you do need a CPT code.

DR. NAEGLE: You need both.

DR. PATING: You need both.

DR. MARK: That's not what's clear to me. That's what we could use clarification on. What CPT codes do you use in that?

CO-CHAIR PINCUS: It sounds like it's a very broad range of CPT codes.

DR. NAEGLE: Uh-huh.

CO-CHAIR PINCUS: It's a very broad range of CPT codes, with some minor exclusions.

MR. GOTTLICH: Right.

DR. SUSMAN: But is the outpatient visit specified within the mental health sector, or is it more broad, including any outpatient? I just want to hear yes or no. I'm reading what the materials are, but --

DR. PATING: My understanding is that they include mostly therapy type visits. You cannot come back to your internist and say that counts as your second visit. You'd have to have a --

DR. CHALK: What the materials say here, in 2(a)(1), ambulatory care clinic, urgent care and clinician office, also behavioral health, also outpatient, also emergency. But ambulatory care and clinic are not specified to specialty treatment.

DR. PATING: I think we have to dig down into the CPT codes to see what they actually map to. That's where you have --

DR. CARNEY-DOEBBELING: The CPT code doesn't map to a provider type. So on 99401, I pull up a bunch of them. These range from

everything to 60-minute psychotherapy visits to SBIRT visits. All of those codes are included, and it does not specify provider type.

DR. MARK: Yeah, but are they all psychotherapy codes? I mean I wish --

DR. CARNEY-DOEBBELING: No, they're not. They're internist or pediatrician codes, the screening and brief interventions.

DR. SUSMAN: So they're E&M codes also?

DR. CARNEY-DOEBBELING: There are HCPCS codes, CPT codes. They're not E&M.

DR. MARK: They're not E&M.

DR. CARNEY-DOEBBELING: They're CPTN.

DR. SUSMAN: So again, if they're not 213, 214, 215 type of codes, then -- and I don't know. I'm asking the measure developer. It would seem to then limit the possibility of follow-up at an outpatient internist's office.

DR. CARNEY-DOEBBELING: No, because an internist could bill an SBIRT code, which are included.

DR. SUSMAN: But if that's the way you're defining the follow-up, by billing that code, I would argue very strongly that most internists, most family docs, most primary care wouldn't be using those codes routinely. There's a much higher likelihood of using E&M codes, which certainly --

DR. CARNEY-DOEBBELING: I'm not arguing one way or the other.

DR. SUSMAN: Yeah, no. I just want to know what the facts are in the --

DR. CARNEY-DOEBBELING: So I think, David, living in the world of HEDIS and NCQA a lot of the last six years of my life, I can argue both sides of this fence.

But I'm going to get into David's head for a minute, because I think your essential problem with ICD-9 codes that are included here is that they include both substance abuse and substance dependence.

So I'm thinking that maybe you're believing that these are too broad, because does

everyone with substance abuse benefit or need the same kind of treatment that substance dependence does, and that the broad range of coding that's allowed in the treatment, to your point, may not always catch what's actually being done in the real world, because internists or general practitioners don't know what SBIRT codes are, or they won't be coding psychotherapy.

So it's forcing some specialty care in a way, because they won't get paid, or else the visits may be happening but they're not being picked up, because they're not being coded. So it presents all sorts of messy measurement --

DR. SUSMAN: That's the issue.

CO-CHAIR PINCUS: So I'd like to summarize. So the argument is -- and then let's hear from the measure developer and then let's vote.

So the argument is that the reliability is potentially undermined by variations in coding practices that may not

capture some of the -- in a consistent way across plans and providers -- the delivery of services that were intended to be within the numerator?

DR. NAEGLE: Yes.

CO-CHAIR PINCUS: And the question is, how extensive is that, and to what extent does that undermine the usefulness of this?

DR. PATING: And just so I can be clear, so the VA data from the 2011 Harris system, in the non-substance abuse clinics, the range for facilities was anywhere from 18 percent to 68 percent of the visits that were coded positive, that they met the criteria when they actually did the chart review?

There was no visit there that really met what was the intent of the coding. So there's this disconnect between the CPT, the diagnosis and what actually happens, and again, this from the VA system data. So that's where this apples-to-apples problem comes in.

DR. CHALK: And I think that's probably true for everything else we've been

talking about today, or will be talking about, that we are a state in this country right now where that kind of complicated problem exists, and is going to have to be sorted out over time.

CO-CHAIR PINCUS: Tami, and then let's vote, and Madeline, you have the last word. First Tami.

DR. MARK: So just two questions for David. Is it your sense that more substance abuse treatment is being provided than is being coded, and secondly, do you think if we include folks who get a diagnosis of substance abuse and they're put into treatment, as defined here, there will be some kind of negative consequence, or it would be unnecessary?

DR. PATING: Again, I think with regards to the patient care aspects of this, there's good things that are happening. I like the intent of this. But because it's messy and being used as a public accountability measure, it has complications.

I would hope that NCQA would continue

to refine and work on this measure. It's going to get worse with DSM-V comes on out, in which we have no more abuse and dependence.

We're just going to have substance use disorder, zero to ten. What is that? Do you just like -- does everyone then get -- I don't know what a zero to ten is. But that's what the new DSM-V will be.

So somewhere in the middle, we're going to need to set some thresholds, and I think that, you know, it's going to get sloppier, and you're going to be looking at measures where people are 15 percent screening, and they'll say they're doing poorly and the other systems are doing great.

But we don't really know with this indicator where people really are, and whether they're getting to where they need to get. That's kind of my point on that.

CO-CHAIR PINCUS: Madeline.

DR. NAEGLE: David, I just wanted to say that I think your points are excellent. We

didn't have a chance also to talk about the fact that a lot of the evidence is out of the VA, which I think in certain ways is a biased sample, in relation to trying to identify a measure, gross though it is, for getting people access to care.

But we really have very little information on populations that we would consider to be in groups where disparities exist, and of course access is a very big issue in those groups.

So I have concerns about the reliability, but I also tend to move in support of Mady's comment, that we don't have anything, and if you predict that we're going to be in worse shape when we get to DSM-V, you know, I think there is an important role for trying to get some assessment of the extent to which people who are identified are gotten into a system of care. So that would be my final comment.

CO-CHAIR PINCUS: Okay. Are we ready to vote? Nancy, you had something up before, but has that been covered?

DR. HANRAHAN: I just wanted to say that the CPT codes that are used there, the HCPCS, all the codes, cover everything. It covers everything, you know, and it seems to me that the intent is to try to cover everything, so that we can pick it up, and yes, that's the best we can do right now. That's all I have to say.

CO-CHAIR PINCUS: Okay. Let's vote.

MR. WILLIAMSON: We will now vote on the reliability. Begin voting now. High, moderate, low, insufficient. We have 10 moderate, 7 low and 2 insufficient.

CO-CHAIR PINCUS: So it squeaks by. Okay. Let's move on to validity.

DR. NAEGLE: Okay. So validity, the initiation and engagement of alcohol and other drug dependence measures, was tested for face validity, expert panels, looking at the measures being consistent with overall performance measures. That material is extensive in the review.

I would raise some question regarding the points that we actually just discussed under reliability. But validity is being questioned from the perspective of diagnostic categories and groups.

CO-CHAIR PINCUS: Other comments on validity?

(No response.)

CO-CHAIR PINCUS: I'd like to step out of the chair for a minute and make a comment on validity. Just when this was initially proposed, I recall -- and Mady, you may want to correct me -- but it was part of a suite of three measures, identification, initiation and engagement.

My own view is that by leaving out identification, it creates a problem, because -- -- and this came up in a direct way, in a study we did evaluating the quality of care at the VA, and comparing it to private health plans.

Because if you have an intensive screening program, you're going to identify a

much larger, broader, more heterogenous group of less motivated people.

So your performance on the initiation of engagement is going to look worse at a place that essentially only gets people that sort of are -- really want to get treatment, and ignores everybody else.

So for example, so when we did this in the VA, the VA performed better on most of the other measures across mental health, with the exception of this one, compared to private plans. But they also had a 200 percent greater identification rate, and so --

DR. CHALK: You're going to have to talk into the mic.

CO-CHAIR PINCUS: So I'm just -- that does have a role in affecting the validity of this, in terms of accountability measures, and so one thing I would say is, it would make a big difference if you also include an identification measure with this.

DR. CHALK: I would comment that I

don't entirely disagree with you. That was not a decision the Washington Circle made, to merge, as it were, identification and initiation. That was a decision made by the developers, who may need to respond to that. But I do think it matters.

CO-CHAIR PINCUS: Jeffrey.

DR. SAMET: More of a question. I think your point's really interesting, and the data bears it out. What does that do? I mean, how harmful is that to the validity, as a consequence of what you said?

CO-CHAIR PINCUS: I don't know. I mean it does raise a question, and it's actually readily fixable, because you have the data.

You know, you could make it a tripartite measure, because you have to have the data in order to produce the initiation measure. So it's not, you know, there's no extra work involved.

DR. NAEGLE: So do we want the measure developers to speak to this?

CO-CHAIR PINCUS: They seem to be pondering it.

DR. CHALK: No, actually I'm not sure that I could agree with you that the data's already there, because I think that health systems that screen for alcohol and other drug misuse, you know, might do it. But that's not actually what this is looking for. This is looking for people who have a diagnosis --

CO-CHAIR PINCUS: Right. I'm not talking about screening. No, I'm saying the identification measure is people who had a single visit.

MS. BARTON: Identification measure.

CO-CHAIR PINCUS: That was the definition of the identification measure. So you have to have, you have to identify the people at a single visit meeting the criteria for ICD and CPT and HCPCS, whatever it is.

MS. BARTON: Well, then you're right. That would be the denominator of the current initiation rate.

CO-CHAIR PINCUS: Right, right. So if that was a reportable measure, you have the three lined up and you could then say that -- in a way, it's similar to something we're going to talk about tomorrow, which is the schizophrenia-anti-psychotic measure.

Having one prescription for an anti-psychotic is not terribly meaningful. But linking it to a medication possession ratio is important. And so this is the same kind of concept.

DR. PATING: Yes. So like within our Kaiser California system, we have parity level coverage and near universal, either same day or next day access, and we are up to about 50 percent screening across our systems, and we are picking up everybody, you know, that has problems and referring them on.

But I think the problem is really when you compare like the private health systems, where it's all done by CPT billing codes. You don't really -- you're not going to send in a bill

unless you're going to do something with that person and send them somewhere.

So I do think that people in the private sector are not diagnosing until, or post, after they make the referral.

(Off record comments.)

CO-CHAIR PINCUS: --associated with higher enrollment.

DR. PATING: Yes, with higher enrollment.

CO-CHAIR PINCUS: With higher enrollment, as compared to how many people get that for one visit. So basically, you know, you're looking at one to one and a half percent in most private plans, as compared to -- actually, it's a lot more than 200 percent, as compared to 20-23 percent in the VA.

DR. PATING: Right.

DR. BURSTIN: Just one comment, of course. You can only evaluate the measure before you. These are good suggestions, of course. The other thing is I'll point out: this

measure's been retooled for EHRs for meaningful use.

So there are real opportunities, I think, to take some these suggestions and build them into the measure we really would, I think, prefer, have it not be based on claims, but in fact be based on good clinical quality data. So I think for good food for thought for NCQA.

DR. NAEGLE: Helen, in thinking about how it might be improved, do you think that we could also highlight for the future the effort to gather more data on disparities in minority groups, vulnerable groups who are certainly represented by plans? And it doesn't seem that we have any information that supports our initiatives in that direction.

MS. FRANKLIN: Excellent point.

CO-CHAIR PINCUS: Excellent point. Other comments on validity?

(No response.)

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

MR. WILLIAMSON: We will now vote on

the validity. Begin voting now.

CO-CHAIR PINCUS: So let's move on to usability.

DR. NAEGLE: How did -- what was your vote there? I didn't hear it.

CO-CHAIR PINCUS: Oh, the vote was on --

MR. WILLIAMSON: Mostly moderate.

CO-CHAIR PINCUS: Yeah. Moderate had 13. I don't know what the other was.

MR. WILLIAMSON: We had 0 high, 13 moderate, 3 low and 3 insufficient.

DR. NAEGLE: Thank you. I must say, I've never felt quite so disadvantaged being on the other end of the phone. I wish I could be there. So looking at the feasibility component, clearly the purpose is quality improvement. It's meant for public reporting. It's useful for public reporting.

I think given the points that have been made about some of the other limitations, I would suggest that it's moderately useful for public

reporting.

Certainly, the data would be useful in quality improvement, and I felt -- given the discussion and the discussion within our small work group -- that I would support this for it being, the criteria being moderately met for usability.

CO-CHAIR PINCUS: Other comments on usability?

(No response.)

CO-CHAIR PINCUS: Ready to vote.

MR. WILLIAMSON: We will now vote on --

CO-CHAIR PINCUS: Oh wait, Jeffrey.

DR. SAMET: Just as another argument, this is where I thought it was best at, you know. It was -- its best quality was usability. So I would be more enthusiastic, and anyway --

MR. WILLIAMSON: We will now vote on usability, and this is a high, moderate, low, insufficient, 1-2-3-4 question. Begin voting now. We're waiting on one response. For

usability, we have 3 high, 13 moderate, 3 low and 0 insufficient.

CO-CHAIR PINCUS: And so now we move on to vote on endorsement overall, and are there -- oh feasibility, feasibility.

DR. NAEGLE: Feasibility, yeah. Some of the difficulties around data generation certainly throw feasibility into question.

Electronic resources or sources, the HEDIS system has established and available. The indications on that I guess also tie in to limitations within the overall HEDIS system across the country.

Susceptibility to inaccuracies, errors or unintended consequences, I think provide a burden and provider knowledge about the use of a range of coding in relation to substance use disorders could potentially be problematic.

The data collection strategy, I think, is certainly moderate to high. So my vote on feasibility, I would support a moderate on

feasibility.

CO-CHAIR PINCUS: Other questions or comments on feasibility? David.

DR. EINZIG: So assuming a person has an accurate diagnosis, then one would assume that there should be adequate access to care. Unlike California and where I come from in Minnesota, that's the not the case.

So frequent follow-up, as frequently as they get into some clinic that quickly and to have that much, that many follow-up visits, that's going to be the challenge, and I assume in other places of the country too.

DR. NAEGLE: Uh-huh.

DR. SUSMAN: I mean, isn't that the point here? I mean we're trying to drive access through accountability at a health plan level. I mean it seems like that's the purpose of this measure, even though on the ground, I hear you, and certainly there is wide variation.

But that's really why we measure and show differences, and hopefully you're going to

help minimize those disparities.

DR. EINZIG: So we're not talking about access to care. We're just talking about recommendations that this should have it.

CO-CHAIR PINCUS: It's the feasibility of actually being able to measure it.

DR. CHALK: If, to respond to your comment, if we were talking about access within a week of initiating, you know, identification and initiation and engagement, I might agree with you. When we have a space of 30 days, which is I think what we're talking about here, I think we need to drive quality.

DR. NAEGLE: Uh-huh.

CO-CHAIR PINCUS: Okay. Any other comments? Okay. Let's vote on feasibility.

DR. NAEGLE: I think -- excellent comment.

CO-CHAIR PINCUS: Let's vote on feasibility.

MR. WILLIAMSON: We will now vote on

feasibility. This is high, moderate, low, insufficient. Begin now.

MR. WILLIAMSON: We've got one more.

MS. FRANKLIN: If we can vote one more time.

MR. WILLIAMSON: Is everyone here?

MR. WILLIAMSON: Is anybody's blinking four red lights?

MS. FRANKLIN: Let's all try one more time.

MR. WILLIAMSON: It's 2 high, 15 moderate and 2 low. That came through.

CO-CHAIR PINCUS: Okay, good. So let's move on to overall endorsement. Does anybody have any further comments or discussion points with regard to overall endorsement? Okay, let's vote. Oh, did you have an overall comment?

MR. WILLIAMSON: We will now vote on

the overall suitability for endorsement. This a yes/no question, 1 or 2. Begin voting now. We'll have everybody try one more time. There we go. All right. The measure passes, 14 yes, 5 no.

DR. SAMET: So before we move to the next protocol, and this may be out of order, and you can tell me to be quiet, but the previous one, not the one we just did, but the one before that, I've gotta admit I thought the last vote was a little bizarre. I'll just give you my take on what --

(Off record comments.)

DR. SAMET: Well, the last vote that took it out of commission.

(Off record comments.)

DR. SAMET: Yeah, it was the reliability. Yes. It was reliability, or it was --

PARTICIPANT: No, it was 1661.

DR. SAMET: Well, whatever it was. It was the alcohol screening, and it was on

reliability.

PARTICIPANT: Yes, it was.

DR. SAMET: Okay, so agreement with that. But as I saw what happened -- which is why I thought it was a little weird -- was that there was this fairly reliable or AUDIT-3 that was used, and then there was an NIAAA, also fairly reliable, one question followed by an AUDIT-C actually, and they're both actually, to my mind, they're both reliable. But what we heard was that they had done the first one, and then when they went and looked at the one question followed by the three questions, because they hadn't gone -- they had used that, they hadn't gone back and revalidated that, it rose in people's mind what we had talked about in the smoking, and people said they never went back and did due diligence, sort of to show that the new one was also effective.

Now I may be reading this wrong, but to me, it wasn't the same situation as smoking, because actually, if they had never taken up the

NIAAA single question one, the original one was fine. That wasn't the same thing as what was in the smoking.

And so it was kind of like it was fine, it was fine. You put it together and people said no, it was insufficient. I thought it had been, I thought there were things that were misconstrued that got that kicked out of the ball park.

So you know, I know it's after the fact. It may not even be proper to bring it up now, but that's how I saw the whole thing, and I thought like -- that wasn't the right way for that one to end, and I was the one that almost had it ending the previous thing.

So I have no real vested interest. I just thought that step of it, that step of it really like kicked it out for reasons that I thought were bizarre.

CO-CHAIR PINCUS: Jeff, it's perfectly appropriate for you to bring that up, and but actually my recollection is somewhat

different, that the issue wasn't the reliability of the instrument.

The issue was the reliability of the measure, and that the reliability, as assessed under the initial testing, wasn't very good, and there were modifications made, and the current version of the measure has not yet been tested. Is that --

MS. WATT: That is not how I would characterize it, but of course, I do have a bias. The measure is the same. What we did was, we tweaked some of the specifications, based on the findings of our reliability visits and the feedback we received from the people who were testing the measure.

CO-CHAIR PINCUS: I think that's the same thing I just said.

MS. WATT: Okay.

CO-CHAIR PINCUS: That the measure specifications were changed, and they had not yet been tested with the changed specifications.

MS. WATT: They are currently in the

process of being used with the revised specifications, yes.

DR. SAMET: For me, I would go back on it. If they hadn't gone back and tried to make it better, where the first set of measures and those, that reliability, I agree, reliability, to document, would that have been sufficient in and of itself?

CO-CHAIR PINCUS: I can't interpret what people's voting was. So, yeah.

DR. SAMET: Okay. I mean, I hear that, but as opposed to what we saw with smoking before, where I would say it was clearly no, here I might argue it was much more in an acceptable range. But you know, if I misconstrued that, then that's fine. But I wanted to at least put that out there.

DR. BURSTIN: Let me just try one thing. I think what Jeffrey's saying, correct me if I'm wrong, is that it seemed like two issues kind of were coming together on that last discussion of reliability.

The first was, was the overall testing data submitted for that particular measure, and I think the reliability there was significantly better --

(Simultaneous speaking.)

DR. BURSTIN: I thought it as higher, sorry.

(Off record comments.)

DR. BURSTIN: Right, and then the second issue was -- then, there was this whole discussion of well, the measure's still being revised. How can we take a measure that's still being revised that hasn't yet been retested?

So I think what Jeffrey's saying was, was the initial reliability estimate sufficient, that perhaps that might have -- if people had separated those issues out, and were people kind of pulling those issues together in their mind as they were voting.

So if anybody else shares Jeffrey's confusion, if it would be useful, you know, we are certainly happy to do a revote, since we need

to go back to those measures anyway.

DR. CHALK: Supposing they hadn't told us, supposing the developers said: well, here it is. It was 252, .252, that's it -- and never said a word about: oh, and by the way, we're respecifying this, because we think we can improve it? Would it have gone through?

(Off record comments.)

DR. CHALK: It's low, but the previous one --

CO-CHAIR BRISS: They and others were inclined to try to tweak what they were doing to make it better, right?

DR. CHALK: Right, and but the other ones, tobacco, was .05. Okay, you have a .05. That's no question about it. .25 can begin to move in the right direction.

CO-CHAIR BRISS: Yeah. We've seen kappas today that range all the way from essentially zero to quite good agreement, that's like on the order of .8. So in the .2 range probably doesn't make me feel so squeamish. If

the baseline had been .6, I'd be right with you.

(Off record comments.)

DR. SUSMAN: Is there any harm in revoting on it? I mean, you know, in a sense, it makes sure that we do our due diligence. I'm not advocating strongly for that. I, in my mind, had it clear what I was voting on, but you know, certainly other people might not have.

Measure 1661 Revisited

CO-CHAIR PINCUS: Seeing none, let's revote. So just to be specific: yes, let's be specific about what it is we're voting for.

This is Measure 1661, which is the alcohol use screening measure from the Joint Commission, and we're voting specifically with regard to the reliability, and we're voting either high, moderate, low or insufficient evidence.

MR. WILLIAMSON: Correct. We will now vote on the reliability. Please begin now.

(Laughter.)

MR. WILLIAMSON: Okay. We have 0 high, 8 moderate, 7 low and 4 insufficient. Move the --

CO-CHAIR PINCUS: How does the non-endorsement of the first one affect consideration of the next two, next three, excuse me. Could the measure developer maybe respond?

MS. WATT: Sure. As you know, the first one has to do with whether or not patients were screened for alcohol use. The second one and the third one, similar to the tobacco measures, talk about whether or not a brief -- two talks about whether or not a brief intervention was offered and performed.

The third was whether or not they get referrals. I'm correct, right? Okay. So excuse me, received treatment or referral. So to the extent that you need to know who's identified as an alcohol user before 2 and 3 come into play, I mean I guess I -- it's possible if there were another method of identifying those

people, 2 and 3 would still be reasonable measures.

DR. GOPLERUD: I believe that 2 is dependent on 1. Three is independent, because 3 is any patient who receives a substance use diagnosis, inpatient treatment is initiated, or there's a specific discharge referral that's totally independent of whether screening was done.

(Off record comments.)

DR. GOPLERUD: Well, the fourth one, if you pull the wings off of the screening part.

(Laughter.)

CO-CHAIR BRISS: Dave's communication, consulting about another pulling the wings off is the right one either.

DR. GOPLERUD: Okay, all right. Let's see. The measure is a complex measure. it has two groups that are in the denominator. One is those who are identified as risky alcohol users, and the second is patients who have a substance use disorder diagnosis.

So to the extent that the measure requires the identification of people who are risk but non-diagnosed substance or alcohol users, then the fourth one also would fail if there isn't a measure to screen. However, it also has the follow-up of patients who are dependent.

But the measure itself would not, as posed, would not work.

CO-CHAIR PINCUS: So maybe Helen, is it okay to ask them if they, how you guys want to proceed, since you're the measure developers? Do you want us to consider the other ones and go through that, or do you want us to just go to three?

CO-CHAIR BRISS: Or another option, I suppose, could be that we could do -- in some ways, these things make most conceptual sense, at least to this reviewer as a coherent set, and trying to approve them one at a time in, you know, in isolation from each other doesn't make a lot of sense.

I mean the other alternative that we might have is do the sort of abbreviated process that we did before with Tobacco-4, and give you general advice on the rest of the suite of measures, since you're going to have to come back eventually, at least with one, right?

We could bring them back as a set, which might be easier for you and easier for us.

(Off record comments.)

DR. BURSTIN: Given the amount of effort put into all these, I still think it would be useful to go through the exercise of having all you in the room and getting the input on the criteria as we go forward. Are you guys okay with that? JHAC in the corner, yes?

(Off record comments.)

DR. BURSTIN: No --

CO-CHAIR PINCUS: No, no. We would vote, just discuss through it.

MS. WATT: Okay. You know I -- of course, I was going to say this. I'm saying this not for the record, but I guess I am saying it

for the record. You know, we're not into wasting your time, and quite frankly, or really interested in being beat up either.

So if you think that, you know, honestly, if you think that there isn't any way that we can defend these measures then, you know, I would love to go through them.

As pointed out, we do spend a considerable amount of time and money developing these measures, testing the measures and submitting these measures.

DR. SAMET: Well, I'm happy. I'm leading this next one. Yeah, I think there's value going through it. We won't have the same passion, because it's not going to make a difference. But it should be informative.

(Off record comments.)

DR. SAMET: Yeah. No, no, no. The other one will be totally the same passion.

DR. SUSMAN: And I guess the other thing I would like to say at least is that I don't perceive we're trying to beat up on any

organization.

I mean there's a tremendous amount of work, and I think JCAHO, NCQA, the other groups that have taken the time and energy to do all this really important work, are doing a great service for medicine and our people in the U.S.

And that I would hope that even if it's very hard to hear some of the feedback, or if you totally disagree with it even, that it would be viewed in the spirit which I think this group is giving it, which is to just call them like we see them. There's clearly a diversity even in this group, and not simply one sentiment that's consistent.

MS. WATT: Thank you.

CO-CHAIR BRISS: And I wanted to say, I said it aside from the microphone before, but I want to, you know, especially in this context I want to say it in the microphone, that both the science and the practicality of these things is a really hard thing, which is part of the reason that all of us are sitting around the table doing

this.

I think that you ran into some roadblocks today, just because the initial stab at reliability testing didn't give us, give any of us quite what we had hoped. I think that hurdle is very likely to get passed the next time you do another round of reliability testing.

So in the big picture scheme of things, I don't think that you've gotten a lot of negative feedback. I think you ran into a minor, what in the big picture scheme of things is kind of a minor roadblock.

MS. WATT: Thank you.

CO-CHAIR PINCUS: But I do think that, you know, I personally think that the work that the Joint Commission and other measure developers do is actually heroic in the face of the lack of sort of resources made available more broadly, to build on the science of measure development.

I think to actually step out in front and to actually take risks and try to put things

together is really important, and it's unfortunate that there's not a broader range of support and resources for that.

DR. ZIMA: There was one other point, I think, I'd like to let JCH know, and that is that they might have been a little bit more transparent about how they measured reliability.

I mean the level of analysis was the patient, and I think that around the room, more people could understand how they measured reliability than NCQA, looking at the state health plan.

Measure 1663

CO-CHAIR PINCUS: So following Helen's suggestion, let's go to 1663, and Jeffrey, if you could -- well first, let's hear from the measure developers, their comments on 1663, Alcohol Use Brief Intervention, Provider-Offered and Alcohol Use Brief Intervention.

MS. WATT: Okay. This measure

builds, as you know, on the first measure. So the denominator is really looking at the patients that screened positive for unhealthy alcohol use as a result of screening with that validated tool.

And in the numerator, we're looking at those patients, then, who received a brief intervention. This is like the tobacco measure, where patients that refused the brief intervention will flow to the numerator, and in the subset measure, then, you see only those patients that actually received the intervention.

DR. SAMET: Okay. So I feel like we've all been trained, I know I have, on the drill. The impact, I think, is pretty straightforward. We actually hit upon it and voted upon it for the screening piece. Alcohol problems are huge, costly and I'm not going to say anymore.

CO-CHAIR PINCUS: So any comments on the impact? Shall we vote? Oh, Lynn.

DR. WEGNER: If I could ask a question. Is the reason that you excluded 12 to 18 year olds is because we just sort of understood that they would fall into the same guidelines as the adults do, with respect to alcohol? Is that why?

DR. GOPLERUD: It was staying on the side of the scientific evidence, as coming from the U.S. Preventive Services Task Force report. There's no reason to think that they wouldn't be, and in fact there is evidence that it is.

MR. WILLIAMSON: Okay. We will now vote on the impact. This is a high, moderate, low, insufficient rating. You may begin now.

MR. WILLIAMSON: Okay, we have 18 high, 1 moderate, 0 low and 0 insufficient.

DR. SAMET: Okay, moving right along. The performance gap is next, and I think again we've spoken to the fact that there was a performance gap with screening. Since it was there, there's going to be a performance gap with

intervention.

But actually, even if you're aware of the citations that are listed in 1(b)(3), there's those and there's others. But I think it's pretty -- I would characterize it as high. By the way, I wasn't at the phone call, so I'm speaking not representing the committee, although their comments are in there.

PARTICIPANT: Be careful with that.

DR. SAMET: Yeah. I'll make sure I have this around with me, because I wasn't there. So there was -- performance gap, when the committee did meet without me, 3 high to medium, 1 low. But I would have made that 4, anyway. But that's all I have to say.

MR. WILLIAMSON: We will now vote on the performance gap. Oh, do we have a comment?

(Off record comments.)

MR. WILLIAMSON: Okay. Begin voting now. We have 11 high, 8 moderate, 0 low and 0 insufficient.

DR. SAMET: Okay. Evidence. So

here, we began the discussion on brief intervention, or maybe I began the discussion on brief interventions.

But we went back and forth, and we actually covered, so we don't have to go through it again, the fact that there are a couple of reviews of the literature, I'd say.

I don't know if they're exactly meta-analyses, but reviews of the literature, which looked at brief interventions. There's clearly kind of equivocal in the outpatient setting, but there have been a more modest number of studies, but nonetheless really that were reviewed before, that show that they're not all consistent.

We went through that as well. Safe paper, you know, didn't show it, but a lot of them did show it. So the quality varied in those, but there were some strong RCTs blinded amongst them.

So I think the group's reflection of that if you have the outcome when the committee,

when the subcommittee met, for quantity, 3-2-1 for high, medium low; quality, 3-0-3. So quality was mixed, because remember, at least half of those were not RCTs. So there was, that's kind of legit, and consistency was 2-3-1, and not all showed the same positive finding.

Yeah. That said, I think that covered it, in brief. But open for discussion.

CO-CHAIR PINCUS: Discussion with regard to evidence? Any comments, questions?

(No response.)

MR. WILLIAMSON: We will now vote on the evidence. This is a yes, no, insufficient question. Begin voting now. For the evidence, we have 14 yes, 0 no and 5 insufficient.

DR. SAMET: Trying to keep up with myself. Next, scientific acceptability -- reliability. The reliability data here, let me pull this out, showed -- give me a second. Here we go. So I can't find the kappas I wrote down. .54 is my head. .537? Yes, I had .54 in my head. Yes, there it is over here. I had it

down.

So this is actually a much higher kappa score than we've seen for some of the other stuff.

(Off microphone comments.)

DR. SAMET: It's from the what? Help me out here.

PARTICIPANT: Of the 19 cases where there was an abstracter and a reabstracter. So they only looked at 19.

DR. SAMET: Yes, so there were small numbers.

MS. WATT: Are you interested in the confidence intervals. So the score was .537, and because of the low in the confidence interval range, from .179 to .894, so on the upper end.

DR. SAMET: Yes. Let's have some discussion. I have a lot to -- let's open up discussion, because I don't think I'm going to enlighten this situation with my own understanding.

CO-CHAIR PINCUS: Comments,

questions? I guess I have a question. I don't really understand the numbers there. So there were 131 cases, but only 19 wound up in the kappa statistic. Why is that?

MS. LAWLER: Well, from what I understand in talking to our statistician, the kappa score is generated on what the abstract or the original abstracter and the reabstracter both agree is in the denominator. That's how he generates the kappa statistic, and in some of these cases, as you saw described here, there were cases that say the hospital felt should have fallen in the population, and we as reabstracters said no, this is not going to fall into the population. It would be what we would call our Categories Bs.

So that's the difference. I don't know. Steven, are you on the line? I guess not.

OPERATOR: And please press star-1 if you'd like your line opened. Star-1, sir. Just a moment please. And your line is opened.

DR. SCHMALTZ: Thank you. Hello?

CO-CHAIR PINCUS: Yes. The question is explaining why the kappa statistics was based on the 19 cases, rather than 131 cases.

DR. SCHMALTZ: Because based our kappa statistic on those cases that both original and reabstracter agreed were in the denominator. It's just so that we don't put the number that weren't in the measure to kind of warp what the kappa is.

So it's kind of a specialized kappa, specialized to denominator cases.

DR. CARNEY-DOEBBELING: So is this to say then that of the 131 cases, there were only 19 instances in which both reviewers agreed?

DR. SCHMALTZ: That they were in measure population, correct. Because, remember, they had to be screened. They had to be screened, and then I don't have the measure specifications. But I think there's a lot of exclusions too.

CO-CHAIR PINCUS: Jeffrey.

DR. SUSMAN: So a question for the measure developers. Where was the discrepancy between the abstraction and reabstraction of the data, such that there was only a very small number, 19 out of 131, that resulted? I'm having a hard time.

If there was that much disagreement, how could there be much stability in this measure? Maybe I'm just not getting it, because it is a little confusing.

But it said of the 19 cases, where both the original abstracter and reabstracter agreed the case was in the population. That's where -- it might be a different 19, but it's 19. I think that's the confusing part of this.

DR. SCHMALTZ: Well, of those 19, there were four false positives, zero false negatives. But since the sample size is so small, the kappa was kind of sensitive to that.

DR. SUSMAN: Let me ask it a different way. How did we get from 131 to 19? What went on in that change?

DR. SCHMALTZ: Well, the 131 were the cases that were reviewed. Once they go through the algorithm, that decides whether they're in the measure population or not. So after exclusions and everything else, that's how many actually end up in the denominator of the measure.

DR. SUSMAN: But how did the case -- how do you decide what gets reviewed?

DR. SCHMALTZ: No, the cases were randomly sampled for review from each of the hospitals. So those all went in, and then once they go through the measure algorithm that says how many actually came out in that particular measure's population. Because we chose the cases for the measure set for particular measures.

DR. MARK: So I think it's the denominator population, right? So the denominator is the number of hospitalized inpatients, 18 years of age or older, who screen positive for unhealthy alcohol use or alcohol

use disorder?

DR. SCHMALTZ: Correct.

DR. MARK: So, no. That's how you get to the 19, because the 131 is just the charts that they pulled, and then --

(Off microphone comment.)

DR. MARK: But the denominator in the measure is different. So --

CO-CHAIR BRISS: So of the charts that got reviewed, there were bigger numbers of people who had tobacco use than there were of people who had alcohol use.

DR. SAMET: But does it beg the question that if you're only looking at 19 people, did you really assess reliability? I mean that's kind of what we're coming from.

DR. SUSMAN: I mean from a meteorologic standpoint, it seems like you want to have 131 cases, not 131 in a larger population, because it's basically meaningless with 19, to my point of view. I mean again, I don't hold myself to be a biostatistician, and

Steve, if you have comments, I'd certainly appreciate hearing them. But it --

DR. SCHMALTZ: Well, I mean if this were a sample for the measure set, not for specific measures, and to get that large a population for that particular measure, you might not even be able to get that, even if you're looking at a year's worth of data for a particular hospital.

DR. SAMET: You have 20 percent or so that are going to be screening positive. So it's not that uncommon. I mean it strikes, I'm getting a little bit more insight into it now, because I admittedly didn't have much before, it strikes me that it's sort of into the insufficient evidence side of the, you know, of assessing this one. I mean I don't understand how you can only have 19 --

CO-CHAIR PINCUS: It sounds like we've beaten this.

DR. SAMET: Yes, okay.

DR. NAEGLE: Yes.

CO-CHAIR PINCUS: So are there any further comments?

(No response.)

CO-CHAIR PINCUS: Okay. I think we're ready to vote.

MR. WILLIAMSON: We will now vote on the reliability. This is high, moderate, low and insufficient. Begin voting now.

Okay. We have 0 high, 3 moderate, 1 low and 15 insufficient evidence.

DR. SCHMALTZ: Let me make one more comment, that the 19 were both agreed that they were in the denominator. But for this particular measure, there was quite a bit of disagreement about whether they were even in the denominator or not.

So for instance, of the original cases, there were 39 where the original viewer put it in the denominator, but only 19 of those where the abstracter agreed that that case was in the denominator.

CO-CHAIR BRISS: So in terms of a lot

of what we're giving is feedback to the developer on this point. So the truth is it looks to me like every time we look at the reliability of the measure, in terms of what we've been able to prove, it looks a little worse, and so --

DR. SCHMALTZ: -- whether they agree it's the denominator or not, and once it's in the denominator, whether they agree it's in the numerator or not. They kind of looked at the second two and not really --

This is a case where it's important to really look at both aspects.

DR. SAMET: Okay. Can we move on?

CO-CHAIR PINCUS: Yes. Let's move onto the next measure.

CO-CHAIR BRISS: This one's kicked on.

DR. SAMET: This was kicked out, you remember? But just to make the quick points on validity and usability, it gets better from here. I mean, validity looked actually -- many of the different measures seem to get high scores

from clarity of the measure, usefulness, interpretability. I'd just give that feedback to you.

And then I actually thought the last couple of things, the usability got high marks on the pilot for .475, .50 and feasibility. You know, just for giving thoughts for the future, the feasibility, it seemed like the future EHR health record, the electronic health record, would be very helpful with sort of the feasibility stuff in the future.

CO-CHAIR PINCUS: Thanks.

(Off microphone comment.)

CO-CHAIR PINCUS: The same 19 cases.

DR. SCHMALTZ: There's more cases in that one.

(Off microphone comments.)

DR. CHALK: Yes. It has more cases, but the kappa went down, compared to the other one.

CO-CHAIR PINCUS: I think we're doing it to give this measure a look, because it is

independent of the other ones. So I think we should review it.

DR. SCHMALTZ: Is that SUB-3 that you're talking about?

CO-CHAIR PINCUS: SUB-4.

DR. SCHMALTZ: Oh, SUB-3.

CO-CHAIR PINCUS: Excuse me, Sub 3, Sub 3.

DR. SCHMALTZ: Okay, SUB-3.

DR. PATING: So Dr. Pincus, a lot of the --

Measure 1664

CO-CHAIR PINCUS: Let's hear from the measure developer. Are you the reviewer, David?

DR. PATING: I'm the reviewer, but a lot of the impact and performance gap has been previously reviewed. There's a little bit of new stuff --

CO-CHAIR PINCUS: Right, but let's hear from the measure developer first, okay.

DR. SCHMALTZ: There are 39 cases that

were included in the --

MS. WATT: Steven, wait just a second. We need to introduce the measure, please.

DR. SCHMALTZ: Okay.

MS. LAWLER: Okay. So as Eric mentioned earlier, this measure is different from the rest of the measures, in that it is independent from the screening. So in the denominator, we're looking at those people who have a diagnosis of alcohol or drug abuse or dependence.

That comes about in two ways, either through the use of ICD-9 CM coding, or if there is explicit documentation by the physician, or another qualified health care individual that has done a further assessment on the patient, that there is indeed an abuse disorder.

So once we've identified these patients for the denominator, then what we want to see is in the numerator is the number of those patients that either were referred for addictions treatment, or given one of the

FDA-approved cessation medications. Also, drugs are involved here.

The other two, the first two measures dealt strictly with alcohol. This deals with alcohol and drugs as well. So drugs enter the picture here. Let me see. Again, we have the refusal component. If someone refuses the prescription or a referral, then they would still flow to the numerator in the subset measure. You see only those that are -- have received the treatment.

CO-CHAIR PINCUS: And just to clarify, this is different from the previously-discussed tobacco measure, and this is referral for counseling or treatment, or medication.

MS. LAWLER: Or. It's an "or" situation; it's not an "and" situation. And it can be that the continued addictions counseling can begin, even where the patient is hospitalized, if someone comes in and does that, or it can be on the outpatient basis, after the

patient leaves.

CO-CHAIR PINCUS: So David, do you want to proceed?

DR. PATING: Okay. Well, I'd like to help us move along quickly. With regards to the rationale and impact for this referral, the developers cite much of the evidence they cited for the previous two studies, and I think that basically we'll leave it at that. So Dr. Pincus, do you want us to --

CO-CHAIR PINCUS: Oh, actually, can we stipulate for the first two, about importance and --

DR. PATING: And gap.

CO-CHAIR PINCUS: And the gap? Okay. Let's move to three.

(Off microphone comment.)

CO-CHAIR PINCUS: Just the evidence, the last one.

DR. PATING: With regards to evidence regarding 1(c), they do add -- the developers cite a different body of evidence, mostly

related to many randomized control trials about the effectiveness of treatment, both treatment as counseling and treatment as medications.

Numerous medications are listed, buprenorphine, naltrexone, what else is in here, antabuse, DoD guidelines recommending all these and project-combined. It's a smattering. Basically, the whole treatment world saying treatment works, and I think it's been -- these are well-accepted, well-standardized and randomized studies.

So I would rate the evidence here in the moderate to moderate-high range with regards to the breadth of effectiveness of referral to treatment. So those are my comments.

CO-CHAIR PINCUS: Are there other comments or issues or anything that people want to address with regard to evidence? Caroline? Oh Jeff.

DR. SAMET: Just there's treatment, and so you say they show how treatment in a number of ways works. Then there's this idea of

referring people to treatment. Is that different, or has that been looked at, you know, whether to refer people to treatment? Should we be thinking about that referral piece?

(Off microphone comment.)

DR. SAMET: David, are you there?

DR. PATING: Umm, yes. I'm looking. I mean I'd turn to Eric and ask if he can comment on that. There are some referral studies. They're kind of varied in between the treatment efficacy studies. So maybe if you could comment.

I think you've got some treatment referral studies in here, as well as many of the studies showing treatment efficacy. Actually, treatment itself works. So I think you've done both here, kind of combining the previous stuff.

DR. GOPLERUD: Because we did have that compound measure, we took a lot of the stuff from NIATx, and what it takes to have a completed referral to increase the rates of completed referral.

But the important number is that 17 percent of people who come in for medical detox or for inpatient services get to an ambulatory or some post-hospital service.

So what this does is really set up the measure of what you're doing is you're either initiating treatment at the point when you identify them in the hospital, or you are sending them somewhere for treatment. By the way, we also make clear that that does not have to be a substance use treatment provider. It has to be just a specific substance use treatment recommendation, which could be from primary care for medication follow-up or a whole variety of things for mental health.

DR. PATING: Would you mind also just commenting, during the measure, you had in the previous conversations we had at the work group, because there was such a high refusal of acceptance of treatment, you developed a submeasure, which was just how many people either got treatment or got medications; is that

correct? That was my recollection.

MS. WATT: I don't know that it's, that there was such a high number of refusals, but those refusals are sitting in the numerator for that first measure. So we just thought, for the sake of public reporting, that it should really be transparent.

When you look at that, that first rate, you don't really know everything that's sitting there. There's people who got treatment and people who didn't. So we wanted to have that second measure, to show that these are -- this is the rate of people who actually received the treatment.

CO-CHAIR PINCUS: Other comments on evidence.

(No response.)

CO-CHAIR PINCUS: Ready to vote?

MR. WILLIAMSON: We will now vote on evidence. This is a yes, no, insufficient vote. Begin voting now.

The evidence, we have 16 yes, 1 no and

1 insufficient.

(Off microphone comment.)

DR. PATING: Okay. So I'm going to actually ask the developers to speak to this data themselves. I think we have the same issues.

We had 131 with regards to the reliability in the sample that was tested, and then there was 39 cases extracted, which were, you know, looked at with regards to the kappa, and they had a kappa of .28.

There seems to be, in looking at those 39 cases, there was some disagreement about what constituted a referral to addiction treatment, and they were not always hooked into the referral system, and it was not quite clear whether they connected or not.

I'm going to actually ask again the developer to explain that point.

MS. WATT: So again, just as we did with the tobacco measures, we required that the referral be made for the patient before he leaves the hospital, he or she leaves the hospital. So

that's what I mean by hooked into the system. What was the other part of the question? I'm sorry.

DR. PATING: Well, that the issue was, you know, can you just talk in the same way we talked about the previous ones, about if you had 131 cases, you extracted 39. You developed a kappa for that, and just speak to the overall reliability of the measure.

MS. WATT: Steven, can you address that please?

DR. SCHMALTZ: Yes. There were -- of the 39, there were five false positives and seven false negatives. Of the original 39 that the original reviewer put in the measure population, 39 ended up -- on whether they were in the population was --

DR. PATING: So the final kappa reported out was .28. To be honest with you, I'm not really sure how to interpret this, whether it's significant or not significant, given our many discussions that we've had.

I think this is a very simple measure. If you look at the actual data elements they pull in, basically was a referral made or not made, and although there's still -- there's a little bit of unclarity about what constitutes and alcohol or drug diagnosis, and we did bring that up in the previous call.

But it seems like something that could be straightened out by looking at correct ICD-9 codes or whatever the hospital codes were. So the actual algorithm was not that complicated. So I'm not sure why the statistic had such a low reliability.

CO-CHAIR PINCUS: So comments, questions? So we have Lisa, Jeffrey and Peter.

DR. SHEA: I just have a question about why NA isn't considered to be a treatment and what the evidence for that was.

DR. GOPLERUD: The National Quality Forum's consensus guidelines are very clear. They say that AA, NA and other support groups are adjuncts to treatment, important adjuncts to

treatment, but are not substitutes for treatment.

DR. SUSMAN: I wanted to try to better understand the issue around referral. Is it simply a matter of actually having physically made the referral and having an appointment set up that's the difference?

So on one hand, it would be having -- you're going to be seeing Dr. Jones on Tuesday, March 15, versus "you need to go see Dr. Jones." Is that the discrepancy or not?

MS. LAWLER: I think that's a pretty good -- oh, I'm sorry. I think that's correct, yes. In other words, we want a referral to be made. We need to know that the referral was made, so the patient knows where to go, and then rather than just saying well, you need to visit, you know, you need to go back to your primary care physician or whatever, we want it to be set up.

DR. SUSMAN: So in other words, if I were to screw up and say "Go see Dr. Jones. Here's his phone number," that wouldn't be

sufficient. But if I made the phone call or had the phone call made to set up the appointment, that would be?

MS. LAWLER: Interesting question. I'm just, I'm mulling that over in my mind for a minute.

So if you wrote the prescription, you gave it to them, I would say no, we would want for that appointment to be set up for the patient, so that he knows that he goes to see someone on such and such a day, or at least there's a referral made to that physician's office or --

DR. SUSMAN: My only concern is if as measure developers, there's some murkiness in what constitutes a referral. I'm concerned about using this, then, for accountability or for even improvement, and there's a fair amount of discrepancy, it looked like, between the initial abstract and the reabstracter.

I'm just wondering how we can reduce that variability. It may be you're already

addressing that, but that's --

MS. LAWLER: But again, I think you've brought up some good points, and things that we certainly will think about, as we move forward.

DR. SUSMAN: Sure.

CO-CHAIR BRISS: It looks to me like, to quote my colleague, I think we could stipulate some of this stuff. I think the themes we've been going through with this measure, that seems to be recurring. So we're talking about small numbers, low kappas, a fair amount of disagreement about what gets into the measure set. And you know, this one looks very much like the last couple to me.

CO-CHAIR PINCUS: Other comments?

DR. PATING: Just one that I thought was a positive of the measure, and I guess we could ask Dr. Goplerud if you can comment. There was a broad range of possibilities of what you would accept as a discharge referral option, what constituted treatment, other than the medications.

Could you describe what you had described to me, what you found would be acceptable? The kinds of treatments, going to a counselor, going to --

DR. GOPLERUD: Yes, sure. I think what we ended up with was in a way very similar to what the NCQA initiation and engagement is, is that the provider is delivering a substance use treatment, and we're not specifying it has to be particular flavor of provider.

So given the importance of primary care in medication-assisted treatment, we wanted to make sure that that was covered, as well as mental health, as well as substance use, but as a treatment provider and not allowing as a number a support group, "You ought to go to AA on Wednesday night."

CO-CHAIR PINCUS: Jeffrey.

DR. SAMET: You know, this has been compared to other ones, but this seems to me a whole lot murkier than the other ones, to be honest. I like the approach that you've taken

that has been brought, but to say that it could be an appointment.

How does someone know if it's an appointment or if some doctor said, encouraged them to do this appointment?

That seems like totally low reliability, just the -- so.

CO-CHAIR PINCUS: Actually, I was going to say that it's something -- let me step out of the chair role for a second. It's a little bit different, than in some ways, you're caught between sort of the reliability and validity issue, between a rock and a hard place, because it sounds like you made efforts to increase the validity by having the measure include that somebody was hooked into treatment as an inpatient, which certainly strengthens the likelihood that it's going to have an impact.

On the other hand, that's led to a greater degree of unreliability, because of the difficulty in ascertaining that. So you know, there's always these kinds of trade-offs in

these things.

MR. WILLIAMSON: Okay. We will now vote on reliability. This is a high, moderate, low, insufficient vote. Please begin now. We're waiting on one more. Is anybody missing? We had one person leave.

(Off microphone comments.)

MR. WILLIAMSON: Yes, yes. We're still missing one then.

MS. FANTA: If everyone could vote one more time. We're still missing one. Maybe it will --

(Off microphone comments.)

MR. WILLIAMSON: Okay. We have 2 moderate, 4 low and 12 insufficient. No, it's a glitch in the system.

(Off microphone comments.)

CO-CHAIR BRISS: --maybe stepping back into a co-chair role, maybe just advice on any of the dimensions that hasn't already been given on one of the other measures.

CO-CHAIR PINCUS: Comments, suggestions, advice?

(Off microphone comments.)

CO-CHAIR PINCUS: Do you want to sort of lead off with suggestions?

DR. MARK: Yes. So this is looking at follow-up of patients, to see whether, to assess whether they're still using, whether they've received treatment, and you know, one issue we had was, you know, and so the evidence that's cited is randomized trials, looking at the fact that follow-up seems to get people in treatment more.

There seems to be good evidence that if you follow-up and call people before an appointment, that gets them into treatment. So it's a little bit of a jump on the evidence, but you know, it's not specific to hospitals and hospitals calling up. But that might be quibble.

One of the issues I had was with the privacy concerns, you know, calling someone

who's been discharged with an alcoholism diagnosis at their home and saying "are you still drinking?" I mean so -- I don't know, Mady. If you want to, or anyone else on the call wanted to bring up any further concerns.

DR. CHALK: Well, yes. It's difficult to know, even though there's a brief, very brief description in there of who would do this, it's very difficult to know where this would be cited, the follow-up after discharge, the continuing care phone calls, and that's a problem, because it's not --

It might be that a hospital would have a capacity to do it, but if not, then this has to be thought through more.

CO-CHAIR PINCUS: Other comments, suggestions? So I guess I would have one. I mean going back to my earlier comment with regard to the tobacco use, you know, I think this would be a very reasonable measure potentially, if some of these performance characteristics work out for an ACO.

For a hospital, it's a but more problematic, as Mady said. But one thing is whether you want it to be just whether there was follow-up to find out what happened, rather than follow-up plus any additional booster of counseling. It seems to me that it's a wasted effort to just call up and say what happened, and then not do anything about it.

DR. CHALK: And what's important about that, though, is that the research, Jim McKay's research and other people's research, indicates that what is important is to do a risk assessment on the phone when you do the follow-up, and to intervene. Otherwise, and it can be done very briefly if you can get to the patient.

CO-CHAIR PINCUS: So it seems to me that would be a much more effective measure, and actually have some impact.

DR. CARNEY-DOEBBELING: And similar to the tobacco, I continue to have concerns about the safety net versus non-safety net type of

hospitals, and working in an inner city setting.

What we see are folks pop in and they get detoxed for a day or two, and they get discharged, never to come back to that hospital system again.

So if something like this goes through, paying attention to what the expected rate should look like, what really is achievable.

CO-CHAIR PINCUS: So you're suggesting maybe possibly risk adjustment?

DR. CARNEY-DOEBBELING: I would love to see risk adjustment, if they can do that.

DR. SHEA: Well, I think it was just to follow up on what you were saying, in terms of not only needing to do something, but then the responsibility and the risk. Where does that fall and for how far out would a hospital need to carry that risk for that patient?

CO-CHAIR BRISS: As with the tobacco measure, you're clearly getting different perspectives around the table.

I'm friendlier, I think, to the issues of when to encourage appropriate follow-up. I think that going forward -- but it's been clear on both of these measures that make the case for that is going to be really important.

So anything you can do to crisp up the case for why this is important for follow-ups I think would help, for hospitals rather, would likely help you.

Another way that I think that you might consider strengthening your case, if you didn't want to try to build in counseling at that point, is to sort of assess whether they've actually followed on the advice they got after discharge, as you did, I think, with the smoking measure.

I think that might be another way to strengthen the case, that this was worth doing.

CO-CHAIR PINCUS: Any further comments or suggestions?

(No response.)

CO-CHAIR PINCUS: Does the measure developer have any comments?

DR. GOPLERUD: We tried to do an awful lot of things in one measure, and I'm surprised that we didn't get, you didn't give us advice about that.

We put together both those people who had received brief advice for risky drinking, and there is some evidence, but there's a new D'Onofrio study that says that doing a booster session doesn't seem to have any more effect than doing a single brief intervention.

But we're -- one part of this, of the denominator are those people who are the risky drinkers, who received a brief intervention. The other part were those people who were dependent, and who are receiving referral for continued treatment.

We're trying to do two fairly different things with perhaps somewhat different populations and throw them into one measure. Was that trying to do too much in one measure?

DR. NAEGLE: Hi, it's Madeline. I

just wanted to reinforce that point, Eric. I think it was something that was alluded in our small work group in discussing that.

But distinctions about the use of interventions with those populations are pretty clear, at least from my understanding of the research literature. So I'm glad that you're making that point. Thank you.

MS. LAWLER: And on behalf of the Joint Commission, I'd just like to thank the Committee for the advice that you've given us today. You've given us a lot to think about as we go back and retest these measures. And again, I'd just like to say thank you very much.

CO-CHAIR BRISS: And I'd like to thank the Committee and the developers and the staff for what I thought was a really productive day, that very nearly got out on time, and I thank everybody for their contributions today.

PARTICIPANT: More comments from the phone, from the public.

NQF Member/Public Comment

CO-CHAIR BRISS: Oh, and I'm sorry. At the end of every session, we need to ask for comments from the phone or from the public. Anybody else like to make comments?

OPERATOR: Again, that is *1 on the telephone.

CO-CHAIR BRISS: Do you have any closing comments?

CO-CHAIR PINCUS: No. I'd like to reiterate Peter's thanks to the staff, the measure developers and to the Committee, and what time are we convening here tomorrow?

Eight for breakfast, and at 8:30 for starting it. Lynn, you have a comment?

DR. WEGNER: I would like to encourage one of our organizations to consider the tobacco screening and apply it to the 12 to 17 population. Teenagers start smoking, they start smoking early, and they dip snuff.

CO-CHAIR PINCUS: Thanks.

(Whereupon, at 5:16 p.m., the meeting was recessed, to reconvene on Wednesday, April

18, 2012 at 8:30 a.m.)