Health and Well-Being Standing Committee: In-Person Meeting Summary
September 12-13, 2016

The Health and Well-Being Standing Committee met in-person on September 12-13, 2016, at the NQF offices in Washington, D.C. to evaluate 24 measures. Additionally, the Committee provided feedback and guidance on the role of the standing committee in overseeing the NQF Health and Well-Being portfolio of measures.

- **Introductions and Disclosures of Interest**
  - The Committee co-chairs opened the meeting with welcoming remarks and staff introduced the NQF Health and Well-Being project team and support staff.
  - NQF's General Counsel reviewed NQF's Conflict of Interest Policy while each Committee member introduced themselves and disclosed any potential conflicts of interest.
- **Project Introduction and Portfolio Overview**
  - Staff reviewed:
    - The current Health and Well Being portfolio of measures.
    - The updates to the measure evaluation process.
    - The role of the Standing Committee in overseeing the NQF portfolio of measures, providing strategic direction for future measure development, and increasing developer involvement in measure evaluation.
    - Ground rules for Committee discussion and interaction, as well as the process for presenting and discussing measures at the meeting and achieving consensus on voting.
    - An overview of the voting criteria and instructions on using the voting software.
- **Overview of eMeasure Evaluation**
  - NQF Staff provided an overview of the evaluation criteria and process for eMeasures, including measures eligible for Trial Use, and fully-specified measures, which included legacy measures and hybrid measures.
- **Measure Evaluation**
  - **Key:** H – High; M – Medium; L – Low; I – Insufficient; IE – Insufficient Evidence with Exception; Y – Yes; N – No
  - **0032: Cervical Cancer Screening (CCS) (National Committee for Quality Assurance)**
    - Developer Representatives: Lindsay Roth, Sepheen Byron
    - Votes
      - Evidence – carried over vote from previous maintenance review.
      - Opportunity for Improvement – **H-1; M-11; L-1; I-0**
      - Reliability – carried over vote from previous maintenance review.
      - Validity – **M-13; L-0; I-0**
      - Feasibility – **H-4; M-9; L-0; I-0**
      - Usability and Use – **H-0; M-11; L-2; I-0**
      - Overall Recommendation – **Y-13; N-0**
    - Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee's deliberations will be compiled and provided in the draft report.

- **0038: Childhood Immunization Status (CIS) (National Committee for Quality Assurance)**
  - Developer Representatives: Mary Barton, Sepheen Byron
  - Votes
    - Evidence – carried over vote from previous maintenance review.
    - Opportunity for Improvement – **H-11; M-2; L-0; I-0**
    - Composite (1d.) – **H-3; M-4; L-4; I-1 (Consensus not reached)**
    - Reliability – carried over vote from previous maintenance review.
    - Validity – carried over vote from previous maintenance review.
    - Composite (2d.) – **H-3; M-7; L-3; I-0**
    - Feasibility – **H-4; M-9; L-0; I-0**
    - Usability and Use – **H-12; M-1; L-0; I-0**
    - Overall Recommendation – **NA**
  - The Standing Committee failed to reach consensus on this measure; specifically, the Committee did not reach consensus on 1d. the quality construct and rationale. It will be designated as "no consensus reached" and put out for public and member comment to gather input from the field before proceeding. The Committee will discuss the measure again on their Post-Comment call held on December 6, 2016. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **0039: Flu Vaccinations for Adults Ages 18 and Older (National Committee for Quality Assurance)**
  - Developer Representatives: Mary Barton, Jenna Williams-Bader
  - Votes
    - Evidence – **H-11; M-1; L-1; I-0**
    - Opportunity for Improvement – **H-11; M-1; L-0; I-0**
    - Reliability – carried over previous vote.
    - Validity – **H-6; M-7; L-0; I-0**
    - Feasibility – **H-6; M-7; L-0; I-0**
    - Usability and Use – **H-6; M-7; L-0; I-0**
    - Overall Recommendation – **Y-12; N-1**
  - Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **0226: Influenza Immunization in the ESRD Population (Facility Level) (Kidney Care Quality Alliance)**
  - Developer Representatives: Lisa McGonigal
  - Votes
    - Evidence – carried over vote from previous maintenance review.
    - Opportunity for Improvement – **H-5; M-7; L-1; I-0**
    - Reliability – **H-1; M-11; L-1; I-1**
    - Validity – **H-0; M-13; L-1; I-0**
    - Feasibility – **H-6; M-8; L-0; I-0**
    - Usability and Use – **H-9; M-5; L-0; I-0**
    - Overall Recommendation – **Y-13; N-1**

- Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.

- **0431: Influenza Vaccination Coverage Among Healthcare Personnel (Centers for Disease Control and Prevention)**
  - Developer Representatives: Megan Lindley
  - Votes
    - Evidence – **H-5; M-9; L-0; I-0**
    - Opportunity for Improvement – **H-3; M-11; L-0; I-0**
    - Reliability – **H-1; M-13; L-0; I-0**
    - Validity – **H-3; M-11; L-0; I-0**
    - Feasibility – **H-2; M-12; L-0; I-0**
    - Usability and Use – **H-11; M-3; L-0; I-0**
    - Overall Recommendation – **Y-14; N-0**
  - Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **0041: Preventive Care and Screening: Influenza Immunization (PCPI Foundation)**
  - Developer Representatives: Elvia Chavarria, Yvette Apura, Diedra Gray, Steven Purcell
  - Votes
    - Evidence – **H-2; M-11; L-0; I-0**
    - Opportunity for Improvement – **H-11; M-3; L-0; I-0**
    - Reliability – **H-6; M-8; L-0; I-0**
    - Validity – **M-13; L-1; I-0**
    - Feasibility – **H-10; M-4; L-0; I-0**
    - Usability and Use – **H-11; M-3; L-0; I-0**
    - Overall Recommendation – **Y-14; N-0**
  - Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3070: Preventive Care and Screening: Influenza Immunization (eMeasure) (PCPI Foundation)**
  - Developer Representatives: Elvia Chavarria, Yvette Apura, Diedra Gray, Steven Purcell
  - Votes
    - Evidence – carried over previous vote.
    - Opportunity for Improvement – **H-10; M-4; L-0; I-0**
    - Reliability – **H-8; M-5; L-0; I-0**
    - Validity – **M-11; L-2; I-0**
    - Feasibility – **H-2; M-10; L-1; I-0**
    - Usability and Use – **H-3; M-11; L-0; I-0**
    - Overall Recommendation – **Y-14; N-0**

- Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.

- **0680: Percent of Residents or Patients Who Were Assessed and Appropriately Given the Seasonal Influenza Vaccination (short stay) (Centers for Medicare and Medicaid Services)**
  - Developer Representatives: Colene Byrne, Laura Smith, Dan Barch
  - Votes
    - Evidence – carried over vote from previous maintenance review.
    - Opportunity for Improvement – **H-11; M-3; L-0; I-0**
    - Reliability – **H-1; M-6; L-;5 I-2 (Consensus not reached)**
    - Validity – **H-1; M-6; L-4; I-3 (Consensus not reached)**
    - Feasibility – **H-13; M-1; L-0; I-0**
    - Usability and Use – **H-12; M-2; L-0; I-0**
    - Overall Recommendation – **NA**
  - The Standing Committee failed to reach consensus on this measure; specifically the Committee did not reach consensus on Reliability and Validity. It will be designated as "no consensus reached" and put out for public and member comment to gather input from the field before proceeding. The Committee will discuss the measure again on their Post-Comment call held on December 6, 2016. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **0681: Percent of Residents Assessed and Appropriately Given the Seasonal Influenza Vaccine (long stay) (Centers for Medicare and Medicaid Services)**
  - Developer Representatives: Amy Helburn, Laura Smith
  - Votes
    - Evidence – carried over vote from previous maintenance review.
    - Opportunity for Improvement **– H-1; M-13; L-0; I-0**
    - Reliability **– H-1; M-9; L-2; I-2**
    - Validity **– H-1; M-13; L-0; I-0**
    - Feasibility **– H-12; M-2; L-0; I-0**
    - Usability and Use **– H-11; M-3; L-0; I-0**
    - Overall Recommendation – **Y-13; N-1**
  - Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **1659: Influenza Immunization (Centers for Medicare and Medicaid Services)**
  - Developer Representatives: Bob Dickerson
  - Votes
    - Evidence – carried over vote from previous maintenance review.
    - Opportunity for Improvement – First vote: **H-0; M-7; L-7; I-0;** Second Vote: **H-0; M-5; L-9; I-0 (Consideration for Inactive Endorsement with Reserve Status)**
    - Reliability – **H-7; M-2; L-4; I-1**
    - Validity – **M-11; L-3; I-0**
    - Feasibility – **H-2; M-8; L-4; I-0**

        - Usability and Use **H-9; M-5; L-0; I-0**
        - Potential for Reserve Status – **Y-14; N-0**
    - Overall, the Committee recommended this measure for Inactive Endorsement with Reserve Status. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3059: One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk (eMeasure) (PCPI Foundation)**
    - Developer Representatives: John Ward, Elizabeth Bostrom, Yvette Apura
    - Votes
        - Evidence – **H-4; M-8; L-1; I-0**
        - Opportunity for Improvement – **H-3, M-7, L-3, I-0**
        - Scientific Acceptability **– H-4; M-8; L-1; I-0**
        - Feasibility **H-1; M-10; L-2; I-0**
        - Usability and Use **H-1; M-8; L-3; I-1**
        - Recommendation for Trial Use – **Y-11; N-2**
    - Overall, the Committee recommended this measure for endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3060: Annual Hepatitis C Virus (HCV) Screening for Patients who are Active Injection Drug Users (eMeasure) (PCPI Foundation)**
    - Developer Representatives: John Ward, Beth Bostrom, Yvette Apura
    - Votes
        - Evidence – **H-0; M-8; L-2; I-1**
        - Opportunity for Improvement – **H-2; M-8; L-0; I-2**
        - Scientific Acceptability – **H-0; M-10; L-3; I-0**
        - Feasibility – **H-0; M-11; L-2; I-0**
        - Usability and Use – **H-0; M-11; L-2; I-0**
        - Recommendation for Trial Use – **Y-11; N-2**
    - Overall, the Committee approved this measure for trial use. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3061: Appropriate Screening Follow-up for Patients Identified with Hepatitis C Virus (HCV) Infection ure) (PCPI Foundation)**
    - Developer Representatives: John Ward, Beth Bostrom, Yvette Apura
    - Votes
        - Evidence – **H-7; M-4; L-2; I-0**
        - Opportunity for Improvement – **H-7; M-5; L-1; I-0**
        - Scientific Acceptability – **H-2; M-9; L-2; I-0**
        - Feasibility – **H-2; M-10; L-1; I-0**
        - Usability and Use – **H-1; M-10; L-2; I-0**
        - Recommendation for Trial Use – **Y-11; N-2**
    - Overall, the Committee approved this measure for trial use. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3071: Follow-up Referral after Positive Developmental Screen (Northwestern University)**
    - Developer Representatives: Ramesh Sachdeva
    - Votes

- Evidence – **H-0; M-2; L-2; I-11**
- Insufficient Evidence with Exception – **Y-10; N-5**
- Opportunity for Improvement – **H-6; M-7; L-2; I-0**
- Reliability – **H-0; M-6; L-7; I-2 (Consensus not reached)**
- Validity – **M-2; L-5; I-7**
- Overall Recommendation – **NA**

- The Standing Committee did not recommend the measure for endorsement; specifically, the measure did not pass Validity. It will be designated as "no consensus reached" and put out for public and member comment to gather input from the field before proceeding. The Committee will discuss the measure again on their Post-Comment call held on December 6, 2016. A summary of the Committee deliberations will be compiled and provided in the draft report.

- **0279: Bacterial Pneumonia Admission Rate (PQI 11) changed to Community Acquired Pneumonia Admission Rate (Agency for Healthcare Research and Quality)**
  - Developer Representatives: Carol Stocks, Cheryl Davies
  - Votes
    - Evidence – did not vote on this criterion. (Measure transferred from current Pulmonary and Critical Care Committee)
    - Opportunity for Improvement – **H-5; M-9; L-0; I-0**
    - Reliability – **H-7; M-7; L-0; I-0**
    - Validity – **M-9; L-5; I-0**
    - Feasibility – **H-11; M-2; L-1; I-0**
    - Usability and Use – **H-3; M-8; L-3; I-0**
    - Overall Recommendation – **Y-12; N-2**
  - Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3067: Human Immunization Virus (HIV) Infection Screening (Centers for Disease Control and Prevention)**
  - Developer Representatives: Abigail Viall
  - Votes
    - Evidence – **H-10; M-5; L-0; I-0**
    - Opportunity for Improvement – **H-12; M-3; L-0; I-0**
    - Reliability – **H-0; M-5; L-5; I-5**
    - Overall Recommendation – **NA**
  - Overall, the Committee did not recommend this measure for endorsement; specifically, the measure did not pass Reliability. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3086: Population Level HIV Viral Load Suppression (Centers for Disease Control and Prevention)**
  - Developer Representatives: Abigail Viall, Irene Hall
  - Votes
    - Evidence – **H-5; M-10; L-0; I-0**
    - Opportunity for Improvement – **H-10; M-5; L-0; I-0**

        - Reliability – **H-0; M-7; L-5; I-3 (Consensus not reached)**
        - Validity – **H-0; M-9; L-3; I-3 (Consensus not reached)**
        - Feasibility – **H-5; M-8; L-2; I-0**
        - Usability and Use – **H-4; M-10; L-1; I-0**
        - Overall Recommendation – **NA**
    - The Standing Committee failed to reach consensus on this measure; specifically, the Committee failed to reach consensus on Reliability and Validity. It will be designated as "no consensus reached" and put out for public and member comment to gather input from the field before proceeding. The Committee will discuss the measure again on their Post-Comment call held on December 6, 2016. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3090: Appropriate Documentation of a Malnutrition Diagnosis (eMeasure) (Academy of Nutrition & Dietetics)**
    - Developer Representatives: Sharon McCauley, Joe Lynch, Angel Valladeres
    - Votes
        - Evidence – **H-0; M-5; L-4; I-7**
        - Overall Recommendation – **NA**
    - Overall, the Committee did not recommend this measure for endorsement; specifically, the measure did not pass Evidence. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3087: Completion of a Malnutrition Screening within 24 hours of Admission (eMeasure) (Academy of Nutrition & Dietetics)**
    - Developer Representatives: Sharon McCauley, Joe Lynch, Angel Valladeres, Alison Steiber
    - Votes
        - Evidence – **H-0; M-8; L-2; I-6 (Consensus not reached)**
        - Opportunity for Improvement – **H-4; M-9; L-2; I-1**
        - Reliability – **H-0; M-11; L-2; I-3**
        - Validity – **H-1; M-9; L-5; I-1**
        - Feasibility – **H-2; M-12; L-2; I-0**
        - Usability and Use – **H-0; M-11; L-5; I-0**
        - Overall Recommendation – **NA**
    - The Standing Committee failed to reach consensus on this measure; specifically, the Committee did not reach consensus on Evidence. It will be designated as "no consensus reached" and put out for public and member comment to gather input from the field before proceeding. The Committee will discuss the measure again on their Post-Comment call held on December 6, 2016. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3088: Completion of Nutrition Assessment (eMeasure) (Academy of Nutrition & Dietetics)**
    - Developer Representatives: Sharon McCauley, Joe Lynch, Angel Valladeres, Alison Steiber
    - Votes
        - Evidence – **H-0; M-8; L-5; I-3 (Consensus not reached)**
        - Opportunity for Improvement – **H-3; M-11; L-1; I-0**

- Reliability – **M-14; L-3; I-0**
- Validity – **M-12; L-3; I-2**
- Feasibility – **H-1; M-15; L-1; I-0**
- Usability and Use – **H-0; M-14; L-3; I-0**
- Overall Recommendation – **NA**

- The Standing Committee failed to reach consensus on this measure; specifically, the Committee failed to reach consensus on Evidence. It will be designated as "no consensus reached" and put out for public and member comment to gather input from the field before proceeding. The Committee will discuss the measure again on their Post-Comment call held on December 6, 2016. A summary of the Committee deliberations will be compiled and provided in the draft report.

- **3089: Nutrition Care Plan for Patients Identified as Malnourished after a Completed Nutrition Assessment (eMeasure) (Academy of Nutrition & Dietetics)**
  - Developer Representatives: Sharon McCauley, Joe Lynch, Angel Valladeres, Alison Steiber
  - Votes
    - Evidence – **H-1; M-14; L-0; I-1**
    - Opportunity for Improvement – **H-1; M-11; L-1; I-2**
    - Reliability – **M-10; L-5; I-0**
    - Validity – **M-9; L-7; I-0 (Consensus not reached)**
    - Feasibility – **H-5; M-9; L-2; I-0**
    - Usability and Use – **H-2; M-11; L-2; I-1**
    - Overall Recommendation – **NA**
  - The Standing Committee failed to reach consensus on this measure; specifically, the Committee did not reach consensus on Validity. It will be designated as "no consensus reached" and put out for public and member comment to gather input from the field before proceeding. The Committee will discuss the measure again on their Post-Comment call held on December 6, 2016. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3039~0421: Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan (Centers for Medicare and Medicaid Services)**
  - Developer Representatives: Anita Somplasky, KeriAnn Wells, Dan Greene
  - Votes
    - Evidence – **H-2; M-10; L-3; I-1**
    - Opportunity for Improvement – **H-8; M-8; L-0; I-0**
    - Reliability – **H-10; M-6; L-0; I-0**
    - Validity – **M-12; L-4; I-0**
    - Feasibility – **H-4; M-11; L-1; I-0**
    - Usability and Use – **H-7; M-8; L-1; I-0**
    - Overall Recommendation – **Y-15; N-1**
  - Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.

- **2828: Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up Plan (eMeasure) (Centers for Medicare and Medicaid Services)**
  - Developer Representatives: Anita Somplasky, KeriAnn Wells, Dan Greene
  - Votes
    - Evidence – **H-2; M-10; L-3; I-1 (**carried over vote from previous maintenance review on companion measure)
    - Opportunity for Improvement – **H-7; M-7; L-1; I-0**
    - Reliability – **H-1; M-13; L-1; I-0**
    - Validity – First Vote: **M-7; L-8; I-0** Second Vote: **M-10; L-5; I-0**
    - Feasibility – **H-6; M-6; L-3; I-0**
    - Usability and Use – **H-4; M-9; L-2; I-0**
    - Overall Recommendation – **Y-14; N-1**
  - Overall, the Committee recommended this measure for continued endorsement. A summary of the Committee deliberations will be compiled and provided in the draft report.
- **3062: Hypertension Screening for Children Who Are Overweight or Obese (Q-METRIC/University of Michigan)**
  - Developer Representatives: Names of Developers
  - Votes
    - Evidence – **H-0; M-0; L-1; I-13**
    - Insufficient Evidence with Exception **– Y-1; N-13**
    - Overall Recommendation – **NA**
  - Note: The developer attempted to reach NQF staff prior to the in-person meeting to withdraw the measure for endorsement consideration. The measure has since been withdrawn.

- **Next Steps/Committee Timeline**
  - Staff reviewed the project timeline and next steps for the Committee. This includes an opportunity for NQF Member and Public Comment on the draft report from October 20-November 18, 2016; a post-commenting conference call on December 6, 2016; which will be followed by an NQF Member Voting Period; a review and final endorsement decision by the Consensus Standards Approval Committee (CSAC); 30-day Appeals period; and submission of the final report to HHS.

NATIONAL QUALITY FORUM

+ + + + +

HEALTH AND WELL-BEING
STANDING COMMITTEE 2015-2017

+ + + + +

MONDAY
SEPTEMBER 12, 2016

+ + + + +

The Standing Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:30 a.m., Thomas McInerny, Chair, presiding.

PRESENT:
THOMAS McINERNY, MD, Chair; Professor of Pediatrics, American Academy of Pediatrics
MICHAEL BAER, MD, Network Medical Director, AmeriHealth Caritas Family of Companies*
RON BIALEK, MPP, CQIA, President, Public Health Foundation
JUAN EMILIO CARRILLO, MD, MPH, Vice President, Community Health, New York-Presbyterian; Associate Professor, Weill Cornell Medical College
BARRY-LEWIS HARRIS, II, MD, Chief Executive Officer, Common Table Health Alliance
CATHERINE HILL, DNP, APRN, Chief Nursing Officer/Director of Quality, Texas Health Resources
PATRICIA McKANE, DVM, MPH, Epidemiologist/SSDI Coordinator, Michigan Department of Community Health

AMY MINNICH, RN, MHSA, Director, Geisinger Health System

JACQUELINE MOLINE, MD, MSc, Vice President and Chair, Department of Occupational Medicine, Epidemiology and Prevention, Northwell Health (formerly North Shore-LIJ Health System)
MARCEL SALIVE, MD, MPH, Medical Officer, National Institute on Aging
KATIE SELLERS, DrPH, CPH, Chief Science and Strategy Officer, Association of State and Territorial Health Officials
JASON SPANGLER, MD, MPH, FACPM, Executive Director, Medical Policy, Amgen, Inc.
MATT STIEFEL, MPA, MS, Senior Director, Center for Population Health, Care Management Institute, Kaiser Permanente
STEVEN TEUTSCH, MD, MPH, Adjunct Professor, Fielding School of Public Health, University of California, Los Angeles; Senior Fellow, Public Health Institute; Senior Fellow, Schaeffer Center, University of Southern California
ARJUN VENKATESH, MD, MBA, Robert Wood Johnson Foundation Clinical Scholar, Yale University School of Medicine

NQF STAFF:
ANN HAMMERSMITH, JD, General Counsel
ELISA MUNTHALI, MPH, Vice President, Quality Management

SHEILA CRAWFORD, Administrative Manager

DIANE FERGUSON, Administrative Assistant

JASON GOLDWATER, MA, MPA, Senior Director

KAREN JOHNSON, MS, Senior Director

ROBYN NISHIMI, PhD, Consultant

YETUNDE OGUNGBEMI, Project Analyst

ALSO PRESENT:

YVETTE APURA, OD, RHIA, PCPI Foundation
DAN BARCH, MS, RTI International*
MARY BARTON, MD, MPP, National Committee for Quality Assurance
ELIZABETH BOSTROM, MPH, PCPI Foundation
COLENE BYRNE, PhD, RTI International
SEPHEEN C. BYRON, MHS, National Committee for Quality Assurance
ELVIA CHAVARRIA, MPH, PCPI Foundation
BOB DICKERSON, MSHSA, RRT, Telligen
DIEDRA GRAY, MPH, PCPI Foundation
AMY HELBURN, PhD, RTI International
MEGAN LINDLEY, MPH, Centers for Disease Control and Prevention*
LISA McGONIGAL, MD, MPH, Kidney Care Quality Alliance*
STEPHEN D. PERSELL, MD, MPH, PCPI Foundation*
LINDSEY MARSHALL ROTH, MPP, National Committee for Quality Assurance
LAURA S. SMITH, PhD, RTI International
JOHN W. WARD, MD, Centers for Disease Control and Prevention
JENNA WILLIAMS-BADER, National Committee for Quality Assurance

*** present by teleconference**

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

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

(8:36 a.m.)

MS. MUNTHALI: Okay, we're going to get started. Operator, Shan, if you can open up the lines and make sure the public knows that we're getting started.

OPERATOR: Your line is live and you are ready to go. Thank you.

MEMBER CARRILLO: Good morning.

MS. MUNTHALI: Good morning. Is that Michael?

MEMBER CARRILLO: This is Emilio Carrillo.

MS. MUNTHALI: Hi, Emilio, how are you?

MEMBER CARRILLO: Hi, good morning.

MS. MUNTHALI: This is Elisa, we're getting started. I just wanted to welcome everyone to the National Consensus Standards for Health and Well-Being Standing Committee Meeting.

This is an in-person meeting for the Health and Well-Being Committee to evaluate and

hopefully recommend the measures for NQF endorsement.

My name is Elisa Munthali, I'm Vice President for Qualify Measurement at the National Quality Forum. And I wanted to just go over a couple of housekeeping items before I hand it over to my colleagues for introductions and to Tom McInerny, who is your co-chair for our opening remarks.

And it's very nice to see everyone. And I wanted to also welcome Steve Teutsch, Matt Stiefel, Barry-Lewis Harris and Anne De Biasi, who's not her today. They're our newest committee members.

So next slide please? Great. For those of you who are in the room, our restrooms are just beyond the elevators to the right. We'll have several breaks during the day. Most of them 15 minutes. We hope, if we make progress, lunch will be longer than 15 minutes and it won't be a working lunch, but we'll see.

In terms of the Wi-Fi, we've listed

the username and password there. So the username is guest, that's lowercase. Password is NQF capital, lowercase guest.

And we ask that you please be in the room as much as possible, your votes are really important. That you mute your phones before we get started. And that if you have to take an important call you just step out of this meeting room.

And just wanted to let you know that we have microphones for everyone. Only three microphones can be on at the same time. If more than that happens you won't be able to speak.

We ask you to move your microphone close to you, as I've done here. Because this meeting is being recorded and transcribed by a court reporter in the back there. And we have folks that are web streaming and also joining us by phone.

So I will now turn it over to my colleague, Yetunde, to introduce herself.

MS. OGUNGBEMI: Good morning. My name

is Yetunde Ogungbemi and I've been with the National Quality Forum for over five years. A little over five years now. I'm excited to work with you all, so welcome. Thank you.

MS. MUNTHALI: And I was going to turn it over to Robyn, she's not sitting there, but Robyn Nishimi is our consultant on the project.

For those of you who are familiar with NQF, she was our first COO. She's worked with us for many years working on disparities and healthcare -- health competency issues. And of course, population health.

And so I'll turn it over to Sheila Crawford for an introduction as well.

MS. CRAWFORD: Good morning. I'm Sheila Crawford, administrative manager for the department. And I'm here to help out without whatever you need me to be. Thank you and welcome.

MS. MUNTHALI: And also helping Sheila is Diane Ferguson who is walking around giving clickers to everyone.

And another colleague of ours, who you'll meet throughout the next two days, Karen Johnson, who is our chief methodologist. She's a senior director here. And she'll be to the left of Yetunde.

So I'll turn it over to Tom for any opening remarks before we turn it over to Ann Hammersmith, our general counsel for disclosure of interests and introductions of the entire committee.

CHAIR McINERNY: Good morning everyone and thank you for taking the time out of your busy schedules to come down to D.C. for this meeting. And we appreciate your time that you also have done in analyzing the measures. And we'll look forward to your comments.

And I also want to thank, very much, the NQF staff, for their very helpful sort of pre-analysis and giving us some ideas about each of the measures. I think it made my studying of the measures significantly easier and I appreciate that very much.

MS. MUNTHALI: Thank you. Ann.

MS. HAMMERSMITH: Good morning everyone, I'm Ann Hammersmith. I'm NQF's General Counsel. I'll lead you through the oral disclosures of interest.

If you recall, we sent you a rather long form where we asked you about your professional activities. And what we do at the beginning of each meeting, the very first meeting, is we ask the committee members to go around the table and tell us if you have anything to disclose.

We are only interested in your disclosure of activities that are directly related to the subject matter before you. So if you did something in another area, you need not disclose that.

We are particularly interested in your grants, research activity, speaking, but only if it's related to the matter at hand.

Just a reminder that you sit on this committee as an individual, you don't represent

your employer or anyone who may have nominated you to sit on this committee.

The other reminder is, some people will, when we go around the table, will say no financial conflict of interest, which is great, but we are also interested in things that may not involve any exchange of funds.

So for example, if you served on a committee that's relevant to the measures before the committee, that may have been without compensation, but we would still be interested in hearing that.

So let's go around the table. If you're on the phone, I will call your name. Tell us who you are, who you're with and if you have anything to disclose. Let's start with the Chair.

CHAIR McINERNY: Hi, Tom McInerny. I am the past president of the American Academy of Pediatrics and Professor Emeritus of the Department of Pediatrics at the University of Rochester Medical Center. I have no disclosures.

MEMBER SALIVE: Good morning. I'm Marcel Salive. I'm a medical officer at the National Institute on Aging of NIH and I have no disclosures.

MEMBER TEUTSCH: I'm Steve Teutsch. I'm mostly retired, but I work at UCLA on economic modeling and at the Public Health Institute in Oakland. And at the USC Institute for Policy and Economics.

I actually work on lots of different metrics group. Do you want me to list what those are?

MS. HAMMERSMITH: I don't think that's necessary, unless you think it's a conflict.

MEMBER TEUTSCH: I don't think it's a conflict.

MS. HAMMERSMITH: Okay.

MEMBER TEUTSCH: But --

(Simultaneous speaking.)

MEMBER BIALEK: Good morning. I'm Ron Bialek, president of the Public Health Foundation. I have nothing to disclose.

MEMBER MOLINE: Good morning. I'm Jacki Moline, I'm the Chair of Occupational Medicine Epidemiology and Prevention, formally Population Health, at Northwell Health in New York. And I have nothing to disclose.

MEMBER McKANE: Hi, I'm Patty McKane and I'm now the Division Director for the Lifecourse Epidemiology & Genomics Division, or actually acting division director, at Michigan Department of Health and Human Services. And I have nothing to disclose.

MEMBER SELLERS: Good morning, I'm Katie Sellers. I am the Vice President for Maternal and Child Health Evaluation and Improvement at the March of Dimes. And I have nothing to disclose.

MEMBER VENKATESH: Hi everyone, my name is Arjun Venkatesh. I'm an emergency physician at Yale University.

I work under contract with the Centers for Medicare and Medicaid Services in development of hospital outcome and efficiency measures. So

none of the measures that are considered here.

And I also do have some research funding from the Emergency Medicine Foundation that uses several of the AHRQ PQI admission rates as an outcome.

MEMBER SPANGLER: Good morning, I'm Jason Spangler. I'm an executive director of U.S. Health Policy and Reimbursement at Amgen. And nothing to disclose.

MEMBER MINNICH: Good morning. My name is Amy Minnich, I'm the Senior Clinical Informaticist at Geisinger Health System. And I have nothing to disclose.

MEMBER HARRIS: Good morning. My name is Barry-Lewis Harris, the CEO of Common Table Health Alliance. And I have nothing to disclose.

MEMBER STIEFEL: Hi, Matt Stiefel from Kaiser Permanente. I'm the director of the Center for Population Health. And I have nothing to disclose.

MEMBER HILL: Hi, I'm Catherine Hill and I am nurse practitioner, a member of the

American Nurses Association, and an inpatient nurse practitioner who still practices daily. And I have nothing to disclose.

MS. HAMMERSMITH: Okay, thank you. Now I'll call on the people on the phone. Juan Emilio Carrillo.

MEMBER CARRILLO: Yes, hi. This is Emilio Carrillo. I am the Vice President for Community Health at New York-Presbyterian. I have nothing to disclose.

Will join you in person later in the afternoon. I will be in travel today. Thank you.

MS. HAMMERSMITH: Okay, thank you. Michael Baer. Is Michael Baer on the phone?

OPERATOR: He has not dialed in yet.

MS. HAMMERSMITH: Okay, thank you, Operator. Before I leave you, I just want to remind you of one more thing.

If during the meeting you believe that you have a conflict of interest, if you think a fellow committee member has a conflict of

interest or if you think someone is behaving in a very biased manner, please speak up in real-time. We don't want you to get three months out and then say, you know, I think I had a conflict of interest.

So if you like, you can speak up in the meeting. If you prefer not to do that, you can go to your co-chair, who will talk to NQF staff, or you can go directly to NQF staff.

Any questions of me or of your fellow committee members? Okay, thank you.

MS. MUNTHALI: Thanks Ann. Next slide. Great. One of the things we forgot to mention is that your other co-chair, Amir, will not be here today but he'll be here tomorrow. There was a scheduling conflict.

So what I wanted to do, before we get started, is to briefly go over our Health and Well-Being portfolio. You heard much of this information when we went through the orientation and Q&A call, but we think it's important just to remind you.

Our work around health and well-being and population health, with regards to the endorsement of measures, started in 2011. With foundational work that one of your colleagues on the committee lead, Steve Teutsch. Steve Teutsch and John Jacobson.

We commissioned them to write a paper and help us to think about how we should be measuring and assessing population health. What we were trying to do is come up with a measurement framework. And as you can image, when you bring so many different stakeholders together, very difficult to land on one of those frameworks.

Ron was a part of that work and so was Jason. And so was Matt. But what we did come up with was really some guiding principles on how we should be looking at population health.

Steve and Don encouraged us not to just look at total population, but also to look at the determinants of health that are outside the clinical delivery system. And also look at

healthy lifestyle behaviors. And also health improvement.

And so our portfolio is really formed around these tenants that Steve and Don worked on and encouraged NQF to follow through. And also on the National Quality Strategies.

So our measures are primarily around health lifestyle behaviors, primary prevention and screening. And we do have a few on those social and economic determinates of health that can be modified. And a number of oral health measures as well.

Right now, today, you'll be reviewing 24 measures. Ten of these have come through our process before. These are what we call maintenance measures. Fourteen are new.

This is pretty unique for NQF. Normally, in a given topical area, most of the measures that we look at are maintenance measures. So this is a good sign that measure development in this area is growing.

The measures are primarily around

primary prevention and screening. A lot of the influenza vaccine measures that you reviewed.

And there is cervical screening, cancer -- cervical cancer screening measures as well. And then there is some healthy lifestyle behavior measures around BMI screening.

Next slide. So this is just a snapshot of the maintenance measures I mentioned earlier. As you can see, most of them are around the influenza vaccines. And we'll talk about this in great detail. There are a number of them. There are eight measures. And then there's one that's an outcome measure.

So those are all process measures. The influenza vaccine measures.

There is one process -- outcome measure, I'm sorry. It's a bacteria and pneumonia measure. That's also a maintenance measure under review.

Next slide. And these are some of the newer measures we have under review. You'll see that many of them are electronic clinical quality

measures, or eMeasures.

That is quite good for this project as well. You don't see that many across our topical areas, but quite a few came in to this project. So we just wanted to give you a snapshot of the new measures that are under review. The 14.

Next slide. So what we did here is breakdown our portfolio so you can really see where we have gaps.

We do have quite a few measures in the primary prevention and screening. A number of, not just influenza vaccine measures, but pneumococcal vaccine measures as well.

But in those measures in which we can see the determinants of health, we don't have that many. And also in the areas of healthy lifestyle behavior or those that are a community-level indicators of health and disease.

Our portfolio is growing around oral health. Last time I think you reviewed about three or four oral health measures. And those were some of the newer measures that came into

our portfolio. So definitely a lot of work in most of these subtopic areas or measure domains.

Do you have any questions? Okay, great. I'll turn it over to Robyn.

DR. NISHIMI: Thanks. Thanks, Elisa. I think I was out for introductions, so I did want to mention one thing.

I was involved in the testing of the influenza measure. So unlike almost all of you, I do have something to disclose. And so I will be recusing myself from all the influenza vaccination measures.

So with that said, can I have the next slide? I'm just going to go over a little bit, to remind you, before you we actually get into the discussion, about the differences in the new maintenance process. Because Tom will be queuing you a few points to ask you whether you feel you need to vote at all, because of the previous committee's deliberations. So I wanted to explain why that would happen.

You all have seen the measure

worksheets. If you log into the SharePoint sites, from your computer here, you'll be able to access the most current ones so that you can see the additional comments that may have come in over the weekend.

There were some more comments this morning, so we'll be populating those. They're largely for measures tomorrow. But if you have a measure tomorrow, you may want to look at that worksheet again.

So on the worksheet, you've seen the preliminary analysis for each measure that was performed by staff. And now you'll see the -- in the orange section, the Committee's comments.

We did not get any pre-meeting or member comments so those aren't there. And then you've had the evidence and testing, if you looked at those. And they're all attached into that big file.

We wanted to emphasize that the staff's preliminary analyses really are just the starting point to help you facilitate your

presentations, as well as pull out, what we thought, as staff, were some of the salient facts. They're just meant as a starting point.

You may completely disagree with them, and that's just fine. And, you know, welcomed. So just wanted to emphasize that we don't intend those to be binding or determinative.

Can I have the next slide? In terms of the way we evaluate the measures, as you know, importance is the first criterion. For new measures, you're asked to look at whether the quantity, quality and consistency of the evidence, whether there were guidelines, whether the evidence was graded, et cetera.

Under the maintenance process there's decreased emphasis here. Developers are asked either to attest that there's no new evidence or update the evidence, if it's relevant.

In many cases, the update is merely updated guidelines, so that the evidence is directionally all still the same. So there's decreased emphasis here and you may decide not to

vote at all.

For instance, when we get to the influenza immunization measures, they all basically have the same measures. So we will queue you at that point to ask whether you just want to vote on evidence for all influenza measures so that you don't have to discuss and vote on it for all eight of the measures. It seems to us to be a little bit ridiculous. So when we get to that point, Tom will queue you there.

The other part of importance is gap. Gap is the opportunity for improvement. With a new measure, we like to see some data, if it's available, either through testing or perhaps the literature in the general area.

But there's increased emphasis on gap, obviously, for a previously endorsed measure. Because hopefully its measures been in use so hopefully there's more gap information.

You may see that the gap isn't closing or there hasn't been any improvement over the

past five years. So then you, as a committee, would have to discuss whether in fact this measure is doing what you would like it to do in quality improvement.

It might be that the mean performance is about the same. But if you look at the range it's been narrowing or there's still a substantial proportion.

You know, the 20th percentile or the 10th percentile for instance are still lagging far behind, so that's a judgment you'll have to make. Whether the gap is still there. But there's increased emphasis on the gap.

Next slide. So for scientific acceptability, there's no difference. You look at the specifications and decide if they're still relevant. Sometimes the specifications have been updated, sometimes they haven't been.

For reliability and validity testing, for new measures, very important that they be tested, very important that they be tested in the data systems that they're intended to use for.

That's an NQF requirement.

There is decreased emphasis on testing of a maintenance measure. Developers have been through the cycle, they've tested it. They may update their scores. They may provide new score level testing where previously they only had data element testing.

But overall, there's a decreased emphasis because the assumption is that the previous committees, again, have looked at this. Have decided that it was reliable and valid.

And so you will be asked whether or not you want to vote on it again. You can always say yes and shouldn't feel shy about saying yes, if you disagreed with the previous committee's assessment.

But overall, there should be decreased emphasis on discussing and picking apart the testing for a maintenance measure.

And then the last two criteria that I just wanted to mention, feasibility. You'll discuss feasibility and vote on feasibility.

There's no difference.

And in fact, with a maintenance measure, they may have uncovered implementation issues. So should vote there.

And then usability and use, there should be increased emphasis. There should be a demonstration now that this is an endorsed measure of use in either accountability or quality improvement. And there should be an identification of whether there is unattended consequences, once it's been implemented.

So again, feasibility, there's no difference, you should discuss whether it's feasible or not. And then usability and use, for a maintenance measure, increase the emphasis.

Any questions on how the difference? And we're going to approach our maintenance first versus new measures. Our first batch is the maintenance, so. Okay.

CHAIR McINERNY: Thank you.

MS. OGUNGBEMI: Good morning. I am going to announce the role of the standing

committee, which is all of you.

So as a standing Committee Member you are to act as a proxy for NQF's multistakeholder membership. You serve in two or three year terms, which is picked at the beginning of every project cycle.

You work with NQF staff to achieve the goals of the project and you review all of the measures in the portfolio, as they are submitted.

You indicate the extent to which each criterion is met and the rationale for its rating, which will be delivered in a report.

You make recommendations to NQF for endorsement of measures. You respond to comments during the review period that are submitted. And you respond to any directions from the CSAC. You also oversee the portfolio of health and well-being measures.

Next slide please. So I'm also going to talk to you about the ground rules for today's meeting. During the discussions, committee members should be prepared with having reviewed

your measures beforehand, which I know that you all did. Thank you very much for doing it on such short notice.

You're basing your evaluation and your recommendations on the measure evaluation criteria and the guidance. And you will use algorithms, which are at each of your places, place seatings, to move through the measures and vote on the criteria individually.

Please remain engaged in the discussion and mute your phones or put them on vibrate, if you are able please.

You can attend -- of course you're supposed to attend the meeting and excuse yourself for breaks if you need them. Keep your comments concise and focused.

And we ask that you please do not repeat your other committee member's sentiments in full. If you'd like to express your agreement, please do so though.

Avoid dominating the discussion with your point. Please make it known what you feel

and we can move on peacefully.

And next slide please. I'm also going to go over the process for measure discussions.

In the back of the room, and on the phone, we have measure developers who are attending our in-person meeting. They will use that time, the two to three minutes that we allot to them, to introduce their measure for two to three minutes.

The lead discussants, which is included on your agenda, they will also introduce the measure to the committee. You'll basically tell the committee, your fellow committee members, anything that you feel is necessary to bring up. Anything that pops out at you or that you would definitely like to discuss with the committee members.

Developers will be available to respond to questions, but only at the discretion of the committee. So you can call on the developer to answer your questions. The developer will not just insert themselves into

your discussion.

Committee members will also vote on criteria and sub-criteria.

MS. MUNTHALI: So we also wanted to go over, very briefly, what we --

MEMBER CARRILLO: May I interrupt for a minute?

MS. MUNTHALI: Sure.

MEMBER CARRILLO: This is Emilio Carrillo.

MS. MUNTHALI: Hi, Emilio.

MEMBER CARRILLO: I will be in transit at 9:50 when 0032: Cervical Cancer Screening comes up and I'm prepared to introduced to that measure. I wonder if someone else from the group, my group, can present it at that time or else change the presentation until tomorrow morning?

MS. MUNTHALI: Unfortunately we have the developers here and they're scheduled to be here. I think Barry-Lewis Harris is your co-discussant. So perhaps, Barry, would you be

comfortable introducing the measure?

MEMBER HARRIS: Sure.

MS. MUNTHALI: Barry said sure.

MEMBER CARRILLO: Thank you very much. My apologies.

MS. MUNTHALI: No problem. So what I'm going to do now is go over voting for the endorsement criteria. This is really just reemphasizing what Robyn went through.

So in terms of importance to measure and report, Robyn told you what is comprised of that. The committee would be voting on evidence.

And if there are no changes to the evidence, you can opt, as a group, not to vote on that. So no change.

This would be only for maintenance measures. Maintenance measures, again, are measures that have come through our process in the past and they're up three year re-look on all of the criteria.

So if you decide not to vote on that, we can skip that. But we definitely want to know

how the measure is performing.

So performance gap is a must vote. You must vote on that. Whether it's a new measure or not.

With regards to testing, the scientific acceptability of the measurement properties, if there are no changes to reliability and validity, the committee can decide not to vote. But if there have been changes, we definitely want you to vote on those criteria.

There are two separate criteria underneath scientific acceptability of the measure properties. Again, if it's a new measure, it's not optional, you have to vote.

In terms of feasibility, with our new emphasis on maintenance to know how measures are doing out in the field, we definitely want to vote on feasibility and usability and use. It's also required.

So next slide. So in terms of achieving consensus. A pass or recommended has

to be greater than 60 percent yes votes of the quorum. So that's not the quorum of the entire committee, it's the quorum of those that are participating in the meeting.

And so for example, 61 percent on yes would be a recommend for endorsement.

Consensus is not reached if it falls within 40 percent or 60 percent, inclusive of the 40 and 60. And of course, that would be a quorum of the committee.

And the measure is not recommended or does not pass if it's less than 40 percent. So 39 percent of votes of yes. And a quorum of 60 percent of the committee.

And I think there are 20 of us here and so we have the numbers. I don't have them off the top of my head, but we have that. So we're going to watch that, make sure we have quorum throughout the next two days.

Okay, great. So as I mentioned before, unlike of a lot of other projects, we have a number of eMeasures in this project. I

think ten.

We do have staff experts with us at NQF that know all of the details and requirements. And one of those is Jason Goldwater, who is our senior director, who specializes in all of our HIT and eMeasures work. And so we thought it would be helpful for him to give a review, an overview, of the measure evaluation criteria for these types of measures. Jason.

MR. GOLDWATER: And thank you, Elisa, for that wonderful introduction that I clearly don't deserve.

So I'm sure this is the part of the morning that you all were looking forward to the most, as we talk about how to review eMeasures right. You don't have to lie, it's fine. If you don't have caffeine, I would strongly recommend getting some soon.

Elisa is right. There were an awful a lot of eMeasures that were submitted as part of this project. Far more than we have seen in

others. Which is great.

This is certainly where we are transitioning to. We're moving ourselves away from exclusively relying on chart extracted measures and into eMeasurement.

Some of you may remember when CMS, back in the good old days when it was called HCFA, when they were first promoting the idea of doing this all those years ago. And it did not succeed then because EHR adoption was extremely low. Both in hospitals and provider settings.

And now, as we are approaching the end of 2016, we're roughly at 80 percent adoption in both provider settings and also in hospitals. So the time for electronic clinical quality measures is certainly here.

And I think it was very encouraging for all of us to see so many health and well-being measures being submitted that were leveraging the data that is found within EHRs, or registers, to be able to look at the performance of a measure without necessarily relying on

manual abstraction.

So this particular project included the evaluation of ten eMeasures. Some of which were being considered for what we call trial use.

So, as I'll get to a little bit later, there's a number of different ways you look at eMeasures. There's really four different pathways that an eMeasure can take as it comes into NQF.

It can be a brand new measures, a de novo measure, it can be a legacy measure, which is a measure that is already in existence in a public program, federal program, that has been respecified as an eMeasure. It could be a respecified measure, which is a claims measure or a chart abstracted measure that's now being respecified electronically, or it can be this, which is trial use.

Why did we come up with trial use? The NQF endorsement criteria for eMeasures requires that when you're testing a measure, it has to be tested in at least more than one EHR

system. Or as I often joke, at least two. It's all in the wording, in how we word it. So at least two different EHR systems.

Now that often leads to the question, well, I've tested it in two hospitals, they both have Epic. Which at this point, who doesn't have Epic. So doesn't that, is that considered two different EHR systems? And yes, it is.

Because there's no one Epic implementation that is like another Epic implementation. They are all very customized for the settings that they are a part of.

But testing a measure in at least two EHR systems, at times, is difficult. Particularly if it's a brand new measure that has never been utilized before.

So the question before NQF was, do we stop that innovation and allow that measure to fall because they are unable to test it, or can we find a pathway that would allow that measure to be implemented because it's clearly filling a gap, it clearly represents innovation and it's

clearly needed, but it hasn't fully met our testing requirements. And so that was the Trial Use Program.

Now the Trial Use Program is a path to endorsement for new innovatively electronically specified measures that can't satisfy our testing criteria yet, but they are ready to be implemented in the real world settings.

The specifications are there, they're able to take the data from an EHR. It is, for the most part, structured data. So it does not pose a burden to collect, but they just simply have not been able to test within two settings.

It's very important to note, very important to note, that when you are looking at eMeasure for trial use, you are not looking to endorse the measure. Now, that's going to be unlike a lot of the things you're going to do over the next day and a half.

Which are, you're going to be reviewing measures and determining whether they should get an NQF endorsement.

When you're looking at a trial use measure, and you're voting, it's not for endorsement. It's to be accepted into the Trial Use Program.

So what does that mean? It means that the measure, if it's approved by all of you, which means you think it is reliable, you think it's valid, you think it's feasible, you think it's usable and you believe that there is enough scientific acceptability to warrant the measure, then the measure is put into the Trial Use Program, which means the developer can go implement it, find places to implement that measure. Whether it's in a hospital or a provider setting.

And they are given a three-year window to implement that measure, collect data. Once they have enough data that satisfies their ability to test reliability, validity, feasibility and usability, then they can come back, to all of you, and say, we've had this measure in the Trial Use Program and here is our

testing data. And then you look at it and determine whether it should be endorsed or not.

So essentially what you're approving is the measure to be tested in a real world setting. Not a simulated setting, or not in a controlled environment in which they would be looking at two EHRs over a defined timeframe, but rather you are approving that the measure can go forward, be implemented.

They can collect data over the next three years, if it takes that long. Once they have enough data, they can evaluate the results, come back to you and determine whether the measure should then be passed and approved for endorsement.

And it is a possibility that when a measure is implemented, the data they get back, may indicate that it's not overly reliable. Or perhaps it is not feasible. That's what the implementation is there to do. Is to test for that.

Next slide. So you will consider the

full NQF criteria when reviewing these measures for approval. So everything that you have already gone over you will look at. You will review these measures as you would any other, it's just simply not being reviewed for endorsement.

Evidence and performance gap, importance to measure and report, R&D voting criteria, as they would be for any measure. So if it doesn't get pass those two, the measure doesn't go forward.

You will vote on one portion of scientific acceptability to determine if the measure specifications are consistent with the evidence. This is a must pass. If it doesn't pass, doesn't get put into the program.

Feasibility and usability and use should also be considered for determining if a measure should receive approval for trial use.

Now you may be asking, well, if they haven't tested the measure, how exactly are we supposed to be evaluating that? Good question.

So what we have allowed the developers to do, when they want to look at a measure for trial use, is to simulate a data set that tests the logic of the measure to determine that the appropriate metric is being calculated as it should be. And the way they use this is through a program through Bonnie.

And I always joke when I say this, because I inevitably get asked this question, well, what does Bonnie stand for? Nothing. I don't know what it stands for.

A long time guess has always been that, having been a developer in a past life, which rarely admit, we are very fond of naming applications after our children or our pets. So Bonnie could be one of those two.

What Bonnie does is it allows you to create a simulated test deck of patients. So patients with criteria and characteristics that would be representative of the real world. And then you can test the measure against that simulated test deck to make sure the measure is

calculating correctly.

Now, when you look at that, it's very important to note that you want to make sure that the measure is working. Which means, it's taking in the patients it should be and it is excluding the patients that should not be put into the measure.

So if somebody presents to you a list of 50 simulated patients and every one of them would fit into the measure, that's not necessarily adequately being comprehensive enough. And that's a question you may want to ask.

They really want to use Bonnie to adequately test, in a simulated environment, whether the measure would work correctly or not.

Next slide. Fully specified eMeasures. The committee will consider the full NQF criteria when reviewing these measures. So every criteria that Elisa went over is the same criteria you would use for an eMeasure.

The specifications are a bit different

than it would be for a chart abstractive measure because you're using coded data within a system. But that does not negate, nor diminishes, the importance of the criteria that you use to determine whether a measure should go forward for endorsement.

The requirements for eMeasure specifications, and I reviewed, I think all of these myself, so I can tell you that most of these at least passed these initial requirements. So an eMeasure has to be in a very specified format.

Without really getting into deep talk, which Elisa and Karen will tell you I do frequently, it has to be laid out in a way that it can be transported from system to system, without the measure or the data being compromised.

So if I am passing my information from one system to another, it looks the same. I open it up and it's the same data that it would be when it was originated. And so that's the called

the Health Quality Measures Format.

It has to be mapped to a data model. In this case we're using the Quality Data Model. And what that means is that the way information is represented, an encounter, a diagnosis, a procedure, is all represented the same way.

So when they represent a procedure, they're using the appropriate elements for a procedure. When they're looking at a diagnosis, they're representing the proper elements for a diagnosis.

And then the measure is populated with that we call value sets. So a value set is basically a representation of a condition, a diagnosis, a procedure.

So major depressive disorder is a value set. And they have coded elements that represent what major depressive disorder is.

Back in the days when this was first started, there was no guidance on value sets. So anybody could create a value set. And that of course led to every measure developer creating a

value set without actually knowing if another measure developer had created something very similar.

So the National Library of Medicine created the Value Set Authority Center, which basically is the library of every available value set.

So when developers are building measures electronically, they go to what we call the VSAC, or this Value Set Authority Center library, and they pull the value sets that they need. If they can't find the value set that is appropriately representing what they want in a measure, then they can go ahead and create one.

What we ask for in all of these criteria, is that the value sets are published. So when you write a value set and submit it to the National Library of Medicine, you have to publish it. And that way everybody can use it. It's not just exclusive to the developer. It's open and available for everybody.

And there have been times when they

haven't been published. So what we want is that the value sets are accessible to everyone, in the hopes that eventually we will see value sets being used repeatedly to represent similar conditions.

The feasibility assessment, which is done through a scorecard, is required to demonstrate how the data elements included in the measure and logic are used to complete the measure. So when they test for feasibility it is, is the logic calculating correctly, does the metric match the objective of the measure, are the denominator and the numerator being populated as they should be, and most importantly, are those patients that should be excluded from the measure being excluded appropriately as well.

Next. So here are the measures that you will be looking at, in terms of those that will be under the trial use, as well as fully specified eMeasures.

I'm not going to spend time reading every single one of these, but certainly you can

see there are quite a bit. Which we find incredibly encouraging. And we find helpful.

Because I think this is, again, is sort of the way where we are transitioning to. And so we appreciate the work on behalf of the developers and the stewards, for moving us in this direction.

What I do want to emphasize, as I often do when I do this, is the criteria that you are using to judge, evaluate these measures, is the same as you would with any other measure. It's not changing.

Except that the testing has to be done in at least two EHR systems. And that when you are looking at trial use, which will be designated by this group, that when you are considering that, you're not considering that for endorsement, it's just simply to be put into the Trial Use Program. And that measure will come back to you, after they have enough data that they can sufficiently test the measure.

Next slide. Okay, any questions?

Sir.

CHAIR McINERNY: I'm wondering how the EHR codes things in the area of the history and the physical exam? And in the history particular, past history, review of systems, family history.

Because some of these measures are very important that the history of whether or not a patient has had in testing or has had a disease, has been recorded. And I'm not sure that the Electronic Health Record is capable of doing that in all cases.

MR. GOLDWATER: That's an excellent question, and you are correct. It is not alike in any one of the EHR system.

In most cases, past family history is what we would call unstructured data. It's a free text that's usually entered into provider notes in the section of the EHR.

Now, there are a couple of EHR systems, that I'm aware of, where you can actually fill out a text field in a dropbox that

would indicate conditions or diagnoses that relate to the family history. Epic has that option, Cerner has that option.

That does not necessarily mean that every EHR does. Nor does it necessarily mean that even when they have tried, to their best to structure that information, that it's giving you the full amount of information that you may deem necessary.

In some of the measures that we initially looked at, that did not seem to present that much of a difficult problem. But that is certainly something to consider in the evaluation.

CHAIR McINERNY: Yes, because I'm reminded of the saying, we see only what we look for. And I think EHR see only what they look for.

And it may be that as we are requiring hospital systems and physicians to do more in the way of eMeasures, the EHRs will need to be structured a little more carefully so that they

are looking for certain parameters to help them decide whether the patient belongs in the denominator or the numerator. And that could be a problem if it's not doing that well.

MR. GOLDWATER: That's correct. And I -- there's only two things I can say. Is that, one, as you approach an eMeasure, that will be described as to how that data was collected. And if it is unstructured free text, then the evaluation will have to be determined whether or not that measure actually is reliable and valid, and more importantly, feasible.

Secondly, half-heartedly, I'm happy to give you the Epic customer service number and you can call and tell them what many of us have been saying for years, that they should find ways of structuring this. Because it does make eMeasure easier.

Particularly in areas such as this. In which past history is a very important component of this.

Any other questions? I must be

getting much more fluent at this, at least I'm assuming. Stunning people in their tracks amazingly.

Okay. Well, I will be around so if there is anything that you all need I'm happy to answer. If I'm not actually present, physically in the office, I'm more than happy to dial in. So if you have any questions, please feel free to ask.

Thank you all very much for your time and I hope you enjoy the next day and a half. I know Yetunde and Elisa will take excellent care of you, as they often do.

MS. MUNTHALI: Thanks, Jason. So we are now ready for review of our first measure. This is measure 0032: Cervical Cancer Screening.

The developers are the NCQA. And we'll ask the NCQA developers to come up to the table, to the left of me.

And while they are preparing to come here, I think they're in-person, we just wanted to remind you about the voting process in front

of you. Everyone should have a blue clicker. Yetunde will go through, after we review the first major criterion, how to use the clickers.

But what we're going to do is have a two to three-minute presentation, by each developer. NCQA will do that. Will discuss the first major criterion on evidence. Then we'll vote on that.

Then we'll discuss, you know, assuming the measure goes through evidence, then we'll discuss performance gap, vote on that.

Then go onto scientific acceptability, vote on reliability. Then have a discussion on validity, and so on. So we'll be queuing you as we go forward. So, Sepheen.

MS. ROTH: Hi, good morning. I'm Lindsey Roth, a senior healthcare analyst at NCQA. And this is Sepheen Byron, Assistant Vice President at NCQA.

So the measure that we're presenting today is a health plan measure. And it assesses cervical cancer screening for woman at the

population level. It's a long standing HEDIS measure that was first introduced in 1993. And the measure was last reviewed by NQF and endorsed in 2012.

And our current measure assesses the proportion of woman ages 21 through 64, who were screened by either cervical cytology in the last three years, or for woman Ages 30 through 64, cervical cytology/HPV co-testing, in the last five years.

And since the last endorsement, we updated the measure to align with the 2012 U.S. Preventative Services Task Force recommendations, in which they had added the recommendation for cytology/HPV co-testing every fives, as an option, in addition to the cytology every three years.

And our data show that there is room for improvement on this measure. There is a quarter of commercial plan members, and a third of Medicaid plan members, who are not receiving the recommended screenings. And there's also

wide variation across health plans.

So for example, there's a 14 percentage point difference between plans in the 10th percentile versus the 90th percentile among commercial plans. And a 27 percentage point difference among Medicaid plans.

MS. MUNTHALI: Great, thank you. So I think for this measure, Barry-Lewis Harris was the lead discussant, and of course Emilio, who's on the phone. Emilio, I'm not sure if you're still with us or you're --

MEMBER CARRILLO: I'm still with you. I'll be chiming in.

MS. MUNTHALI: Okay, great. So, Barry, think it's up to you discuss any points with evidence that you reviewed?

MEMBER HARRIS: So I'll say it's quite interesting this being my first meeting and this being the first measure that I have here. Actually be the first person to talk without anyone else having an opportunity to kind of give me a guide. And then the first name is on the

phone.

So I would like to say that this, my understanding, that this was a maintenance not actually one straight out of the box, is that correct? Because of the --

MEMBER CARRILLO: Barry, right now I'm in a place where I can speak, so why don't you let me get started.

MS. MUNTHALI: Okay, thank you, Emilio. But you're doing great, Barry.

MEMBER HARRIS: Okay. I wanted the baton first.

MEMBER CARRILLO: I'm waiting for a taxi, but I think I can get started. Well, this is a maintenance measure and, as has been pointed out, it's a measure that is primarily a health plan measure. It's a great importance.

Cervical cancer screening has been shown, repeatedly, to be a very, very valuable preventative measure. And it's something that the health plan follows.

And subsequently, we have large amount

of information and can compare commercial plan information from Medicaid governmental information.

It's a maintenance measure, it's a process measure. And it was first updated in 2003. And now we're at the end.

And it was first updated in -- it was first implemented in 2003 and then updated in 2012. And the impact of the update is not relevant.

So I recommend that there's no need to re-discuss the evidence, since that has been done properly before. And this is a maintenance measure.

The important consideration of performance gap for a maintenance mentioned, it has been noted. And we do have evidence of a performance gap. Particularly in terms of commercial versus governmental plan.

The issue that I think is important for us to discuss is that there is no data stratified by race, ethnicity and language. And

this is something that was recommended by the committee in 2012. And is something that should be considered by us.

In terms of some other issues, there is some concern, on my part, about the age ranges. It's not clear to me, from what's being presented, why we have a difference in the denominator age range from the measure age range. And I'd like to get some clarification on that.

In terms of the measure --

MS. MUNTHALI: Emilio? Hi, sorry. We just wanted to talk about evidence right now. I know you mentioned some issues around performance gap and the specifications. And so maybe we can open it up to the committee for any further discussion on evidence only. So then we can have a vote on evidence and then have a discussion on performance gap and the other criterion.

MEMBER CARRILLO: Very well.

MEMBER HARRIS: So I would like to chime in now, again, Barry-Lewis Harris, to say that I think this is, you know, of course was

once one of the most common cancers affecting woman. And now at 14, the number 14, and I think the evidence, again, as he stated, is there that we should definitely keep this in place.

MEMBER TEUTSCH: I agree with all of that. Question though is, there's a ginormous over utilization of cervical cancer screening. And I know that's not directly related to this measure.

But do we have evidence that they can present on how much over screening there is, the consequences and how that might get incorporated?

MS. MUNTHALI: Lindsey or Sepheen?

MS. ROTH: Sure. So we don't have data for woman in this age group with respect to overuse. We do have a separate measure that assess non-recommended cervical cancer screening in adolescents, Ages 16 through 20.

MEMBER TEUTSCH: And fast, can I ask why the high-risk over 65 are not part of this measure?

MS. ROTH: So for this measure, we

have it aligned with the USPSTF recommendations for woman in this age group. And this is for commercial and Medicaid plans. And so we do not have a corresponding measure for adults over 64. And I believe also, the recommendation from the USPSTF is not recommended.

MS. BYRON: Right. Are you asking why there isn't a measure that looks just at the high-risk?

MEMBER TEUTSCH: Right.

MS. BYRON: No.

MEMBER TEUTSCH: Because, I mean you're probably right, because it's not the Medicare plans. And the Medicare plans should also be looking at woman who had abnormal screens, and considered at high-risk, who should have continued screening.

MS. BYRON: Yes. And I'll just also add that specifying in the population, the task force recommendation notes a number of things that would be very difficult to specify, so you would have to be looking for a prior Pap test,

the results, whether or not you were foreign-born from certain countries and that sort of thing.

MEMBER TEUTSCH: I'd just point out that that's true, but the highest risk are the people who ain't never got screened or the woman who had abnormal screens as opposed to people who have been very good about having their routine screening. So it's the group of perhaps greatest interests.

MS. BYRON: I think it's something we can take into consideration. So thank you.

MEMBER BIALEK: Just a quick question. Where in the measure is it specified which plans? So you said it's Medicare and --

MEMBER TEUTSCH: Medicaid.

MEMBER BIALEK: I'm sorry, Medicaid and commercial. Sorry. So where, I haven't found that in the measure itself on the worksheet?

MS. ROTH: So just to clarify, the worksheet includes the measure specifications, but with respect to how it's reported to us,

plans, if they're a commercial plan, would report their data to us and we analyze it. Or if they're a Medicaid plan they report their data to us and analyze it.

And I believe in the worksheet, under the performance data, we have broken out the commercial rates versus the Medicaid rates.

MEMBER BIALEK: Thank you. So just a question for staff. As reviewers, how do we know what the measures are meant to be? Who are they meant to be used for?

Because often, when you look at the measures, the generic measure of X to do Y. But then when you look at the evidence, it's difficult to match the evidence with the population --

DR. NISHIMI: If you look under, on the measure information sheet, there's blue text at the top. And the third cell down says, measure type process, data source and then it says level of analysis. That tells you the level of analysis that the developer is seeking

endorsement for.

So for instance, they're not seeking endorsement for hospital's data. Because they have been tested in data.

Someone else might have a slightly different set of specifications for a measure, but they're only seeking it because they tested it in these systems.

MEMBER BIALEK: Thank you, that's quite helpful. I always looked at level of analysis being the analysis done and the data that were provider, not the specification for the measures and that's really helpful.

DR. NISHIMI: Yes.

MEMBER BIALEK: Thank you.

MEMBER HARRIS: So I have a question related to the same performance gap. When you say the commercial plans were 77 percent and Medicaid plans were 66 percent, are you saying that those particular plans were those who actually were in compliance with actually doing this?

MS. ROTH: So that indicates that is the average performance rate among the plans that reported their data to HEDIS, to NCQA. So across 200 and some plans in the country --

MEMBER CARRILLO: I'm signing off.

MS. ROTH: -- the average is 77 percent for Medicare and 66 percent for the Medicaid plans.

MS. MUNTHALI: Okay, are there any other questions with regards to evidence? And one of the things we forgot to say is that if you'd like to make a comment or have a question, if you can turn up your tents just like this. I think it will be easier for Tom and I to manage.

So because this is a maintenance measure, and as you remember from our intro and today, and the new evidence that NCQA has presented is directionally the same as the evidence that they have provided when it was last reviewed, the committee can opt just accept the evidence, as in the past, and just not vote on it. So wanted to pose that question for you to

see if you wanted to re-vote or just accept the evidence as it was last accepted and the changes that were directionally the same?

And just a yes/no is fine. Or a second. You seconded? Okay. If no one objects to that? Okay, so we'll go to performance gaps. So we will just carry over the votes from last time.

So Barry-Lewis or Emilio, can you lead us in discussion on performance gap?

MEMBER HARRIS: So Emilio has signed off. So we are, just to make sure that I was listening correctly, so we're sort of talking about performance gap in the evidence section.

And one of the questions that was just answered was related to how well the systems were actually doing. And so we've gotten that clarified, at least from my perspective, as far as, is there any work that's being done to try to help or will this measure actually help us close the gap, is that what the intent is for this particular measure?

I know that we were talking about overutilization by Steve a moment ago, but wondering whether or not we were thinking that we needed a greater emphasis on those who were in their appropriate age group. That was one thing I read about the measure only being started at Age 24 versus 21. Is that a gap that we also need to look at?

MS. ROTH: Okay. Yes, so I just wanted to start by clarifying the age group in the measure. So the measure actually assesses whether woman Ages 21 through 64 received cervical cancer screening. But the denominator is specified to start at Age 24.

And this is because there's a three year look at period where we're assessing whether woman had cytology performed between Ages 21 through 24. And we also specify it to start at Age 24 because we don't want to capture the screenings performed before Age 21 since those are not recommended.

So this is really just a measure

specification issue with the measure.

MEMBER HARRIS: And one of the things that was mentioned in the document, or in this particular measure, was related to whether or not there was disparities there as it relates to this. And one of the things I saw that said it, this was a population health, or you were not looking at different types of groups when you actually did this measure, is that right?

MS. ROTH: Correct. Yes. And just to respond to the question about disparities as well. So obviously we do stratify the data by type of insurance, such as commercial versus Medicaid, but we don't currently collect through HEDIS, performance data that's stratified by race, ethnicity and language.

But this doesn't stop health plans from actually stratifying data in that way. And we do know that many healths in fact do do that. We just don't have access to that data.

MS. MUNTHALI: Arjun?

MEMBER VENKATESH: Thanks. I guess I

have one comment, and the other is just an informational question.

The informational question is, that under performance gap in the sheet we were given, for Medicaid it says the mean is 60 percent but the 10th to 90th percentile is 68 to 73. So those numbers, I guess is that the 10th percentile number is wrong?

MS. MUNTHALI: It's a typo.

MEMBER VENKATESH: It's probably 48 percent, looking at the other years, is that right?

MS. MUNTHALI: Yes, that's right.

MEMBER VENKATESH: Okay.

MS. ROTH: Are you -- I believe it's on Page 16?

MEMBER VENKATESH: I'm on Page 3 of 34.

DR. NISHIMI: Yes, the mean is -- for the Medicaid rate?

MEMBER VENKATESH: Medicaid mean, 2000 --

DR. NISHIMI: Is 60 percent.

MEMBER VENKATESH: And then the --

DR. NISHIMI: The 10th is --

MEMBER VENKATESH: -- 10th and 90th is?

DR. NISHIMI: Fifty-four to --

MS. ROTH: It's 46.

DR. NISHIMI: -- 43. Oh, I'm sorry, 46 to 73.

MEMBER VENKATESH: Okay.

DR. NISHIMI: That table.

MEMBER VENKATESH: It's a typo. All right. So my comment is this. I guess, when I look at these numbers, on the commercial side, I see performance rates that have largely not changed over three years. That is not to say that that means the measure is in, you know, automatically flawed, but to me that brings up three questions.

One is, is it possible that because these measures are specified in administrative claims, we don't capture everything we would want

in a perfect measure and this measure is actually already topped out? Meaning, that this is as good as performance as it's going to get and there's not much additional diet in the measure.

Or second, does it mean the measure is not exactly measuring what we want it to? That it's not able to directly measure screening for cervical cancer in the way we intend and that's why you see the same number every year?

And then I guess the sort of related question to that is, this disparity between the Medicaid rates and the commercial rates, is that a disparity?

And if it is, then I think this speaks to why this measure is important and why it needs to remain or is that just simply a function of the fact that there's more turn in Medicaid plans and as a result, they're not capturing at the same rate, data elements that you would capture in the commercial measure, and that's why that also is similarly kind of flat, but just translated down the curve a little bit?

And so the reason I ask these questions is because we're about three years in, we should ask whether or not there is value in continuing the work of the measure, having the measure and doing this measure. And it's possible, three years from now, all these numbers could improve. And we'd look back and it doesn't say, yes, this is worth all the effort and burden. But if not, I'm just wondering if we're just not getting anything out of the mission.

MEMBER SALIVE: I think this one is worth keeping. I mean, as was stated by the primary reviewer, the long-term trend is very positive on, I think, prevention of cervical cancer.

So whatever happened in three years, you know, we can speculate on that. But it still needs to be continued.

MEMBER STIEFEL: I probably only need to ask this once, as a new member. Because I have a feeling it will come up a lot.

So race, ethnicity, language are

obviously important for a measure like this. NCQA and HEDIS don't have access to that information.

So I guess the question, what's the role of NQF here? It looks like it was recommended last time the measure was reviewed.

Do we just have a sort of generic recommendation that would be really good to have race, ethnicity and language data recorded in HEDIS? So I'm not sure where, what the committee's role is here.

MS. MUNTHALI: Yes, this was a topic of discussion last time around. And I'll ask Robyn to chime in as well.

You know, the developer is constrained by what they can get in terms of the data. There was a very strong recommendation that the developer try and find a mechanism to be able to assess across populations and see differences across populations. And it will be good to hear perhaps what the progress has been since then.

MS. BYRON: Yes, thank for you

bringing -- I mean, it is a very important point and we do understand that. And it's something we are exploring for the future in terms of how we can get data.

We do have two measures within HEDIS that look at race, ethnicity, diversity of membership. And another one looking at language diversity of membership.

So it is something that we have looked into for a while. And we've explored how to really encourage and culturally and linguistically appropriate services through other means.

We have a multi-cultural healthcare distinction program that health plans can look at that really layout the standards that you should be meeting to be able to provide culture and linguistically appropriate services.

So that includes things like collecting data for race/ethnicity. Making sure you have network adequacy. You know, making sure that you understand the needs of your population.

Race/ethnicity, those types of disparities, we have heard from health plans, it's very specific to certain regions and areas and plans. So what we're balancing is the importance of whether or not they should be reporting that information to NCQA so that we can show the averages or whether it's something that they should be doing within their own area and slicing the data in the ways that they can, using their enrollment information. So should they look at race/ethnicity?

And then also look at gender, age. And we have found that health plans, because they are motivated to improve their rates, are motivated to look at the data in that way.

That said, we do understand it's important and it is something that we want to explore for HEDIS, and that we're continuing to do.

But through those measures that are in HEDIS now, we can see that a lot of plans, even the ones who are very far ahead of the curve, are

still not reporting that they have complete data around race and ethnicity. And there are a number of ways they can be getting it that they are still exploring.

You know, whether they should be getting it through indirect means, such as zip code analysis, or whether they can get it directly from CMS.

And so those are things that are in the works for us. And we've been working actually with the CMS Office of Minority Health to explore how to continue to do quality and measurement in vulnerable populations. That that includes race/ethnicity, people who have limited English proficiency, sexual minorities and that sort of thing.

MEMBER STIEFEL: Is it appropriate for NQF to make such a recommendation? If it --

MS. MUNTHALI: So it would be the committee doing it on NQF's behalf. So if you think that it is a strong preference or you want to, let's say an annual update at the next

maintenance review, see those data and see the measure at least stratified or looked at by those different sub-populations, that is appropriate.

Steve?

MEMBER TEUTSCH: Two points. One to follow on to what Matt's saying. Maybe the most important, sociodemographic variables, are things like education, income, other kinds of things.

And then in terms of actually improving these measures and allowing plans and other statutes, a lot of them do something about them. Many of those measures should be collected as well. I know that's a reach, and it's not specific to the cervical cancer measurement, but in general we need something richer than race/ethnicity.

And I know that you talked about some of them, like culturally, in appropriate communications and things like that. But it was a much broader set of things that are really important.

To Arjun's point, and someone can

correct me because this is coming out of my memory, there are roughly 4,000 deaths a year from cervical cancer. It's my recollection that's the same number there was in 1980.

And you can correct me if I'm wrong on that, but --

DR. NISHIMI: Deaths.

MEMBER TEUTSCH: Deaths, right? Roughly. Which would suggest that we haven't made a huge amount of improvement.

And I think it's partly because of some of the issues we've been talking about here which is, you know, some people get screened pretty religiously and then there is a cohort, many of whom are actually not in reality, healthy systems that you're measuring, who are not get the necessary services.

So it's not really a reflection of this measure and how much these plans are doing things, but sometimes I think we are looking under the light post when the problem is elsewhere.

MS. MUNTHALI: Patricia? Barry?

(Laughter.)

MEMBER HILL: Oh, I'm happy to speak up, because we talked about this the last time I was here. And I would like to say that the committee, I think, is interested in seeing the measure developers take some lead.

I understand there will be questions, at a systems level, as to who it's most useful to. But somebody has to step out there and take the lead. And we'd love to see the measure developers do that consistently.

I dream of a time when we meet and we don't have to bring up the issues of disparities and parsing our data sufficiently enough to address all American needs in health.

MEMBER HARRIS: I was just going to actually just bring up the additional point that looking at the race/ethnicity is nice as well, but I think looking at the other areas that was mentioned, mostly related to education as well as demographical variances or differences, because

it is 4,120, I think, is the estimated amount.

So it's the same number that he stated, but I think that, you know, are we really looking at the population that needs the greatest impact or are you just looking at the same population that we've always looked at for several years?

MEMBER McKANE: Okay, I was just going to say that with many of our metrics that we look at in public health, often race/ethnicity, we do see racial/ethnic disparities. And when we do call control for economical education, the disparities remain. So I think that that is very, very important.

And I also fully appreciate that with claims data or electronic data, that it can be difficult to capture race and ethnicity accurately. You know, ideally it's a self-report, but it may or may not be. And people have the right to refuse to give that information.

So I think also, I remember one

measure I reviewed they did, you know, there was a 13 percent missing on the race/ethnicity variable, which could skew your results. But I would echo and encourage that that analysis be done.

MS. MUNTHALI: Ron?

MEMBER BIALEK: I'm just wondering if NQF, at some point, can add to its criteria for maintenance measures, presenting data from national data, it could be from BRFSS or other sources, that during that period of time, when the measure was in effect, potential consequences. And it could be unintended consequences.

So you can remain at 4,000, but you can be improving with one population and declining for another. And I think it's difficult for us, in a maintenance measure, to say, go forth with the maintenance measure without having any of the national data of what impact, nationally, or what unintended consequences could there have been as a result of

the measure.

So I don't know if NQF has that in the criteria that the measure developers are to present other data that can be national, generalizable data, to suggest what's going on within that particular health condition.

DR. NISHIMI: That's actually one of the intents of the disparities question. The developer does have the option to present non-performance measure.

The preferred obviously though, as NCQA put it, which is the performance data based on their measure.

A question for them though. You indicated that health plans are using this measure, obviously, and collecting it on their own. Is it possible for you to work with some of your health plan to anonymize and just bring back to the committee some of the disparities findings they are having, you know, perhaps in an annual maintenance? Certainly by the next full maintenance.

Because this is four years now and we still have no disparities data. So I think you can hear a little bit of the frustration from the committee on not having the data.

And if you can either look at the literature or work with your health plans, I think that would probably go a long way to helping the committee out. I see nods around the head.

CHAIR McINERNY: I know that many physicians are reluctant to record race and ethnicity in their electronic health records or their paper records anywhere, and so that must make it very difficult for NCQA to get race/ethnicity data. Do you have, in general, do you know what percent of physicians are recording race and ethnicity and can you at least give us a feeling for that please?

MS. BYRON: Yes. I think it's changing. We looked at a health plan level, but obviously the health plans are pulling from physicians and others.

The last time I looked, but that was a while ago, you know, like I said, the leading health plans who were actually putting a lot of effort into this area, maybe had race/ethnicity data on about 30 percent of their population. This was a little while ago. I think it's probably, hopefully, a little bit better.

But, you know, there are issues with ONC requiring the race/ethnicity variables for meaningful use that I think is going to be helping things to be more standardized. You know, just agreeing on the types of categories one should be using.

That's always been a back and forth because some people want to collect very detailed information based on the area in which they reside. So there has been issues rolling up to the same aggregate categories. So even when you have the data, sometimes it's not very useable.

When we looked at our race/ethnicity variable in HEDIS, we saw that many plans were reporting unknown or refused. And I think

there's probably a lot of that still going on.

MS. MUNTHALI: Other questions? Okay, I think we're -- oops.

DR. NISHIMI: I was just going to say, I think we're ready to vote, but I just want to confirm with the meeting that we do want to make a recommendation to the developer to try and work with some of their members who are perhaps collecting and have a better feel front line. So that the next time you look at this measure you will have some kind of -- so we'll do that for you, okay.

MEMBER HARRIS: So do we need to separately vote it or --

DR. NISHIMI: No. I just wanted to confirm by looking at it.

MS. MUNTHALI: It would go as part of your final recommendation in the report.

Okay, so before we actually go through the vote we'll have Yetunde go through an example or show us how to do it.

MS. OGUNGBEMI: So I'm going to give

you instructions on how to vote. When voting, please use your blue remote control. Everyone should have one. If you do not, please see Sheila or raise your hand and she can give you one.

Only press the voting options that are made available to you. So it will usually be voting options one through four. And sometimes it will be one or two. Sometimes it will only be one, two or three.

Point towards my colleague Sheila and Diane over there beside the windows when you are voting because they have the software to capture your votes. If you change your vote while the voting is still open, it will only capture your last vote.

So if you press 1 and you meant 2, you can press 2, while the voting is still open. If you have any technical difficulties, please raise your hand immediately so we can attend to you then.

I will be proxy voting for Michael

Baer and Emilio, once he joins us again. Those are the participants of the committee that are on the phone. They will be submitting their votes confidentially via the chat option on the web platform.

And if you have any questions, please let me know. Okay.

MS. MUNTHALI: So we're going to be voting on performance gap. Oh, you want to do a test? Let's do a test. So we'll do a test on evidence. We're going to do a test on evidence.

As you remember, we decided not to re-vote but we just want to see, make sure you captured the instructions and your clickers are working. So I will turn it over to -- Sheila, are you going to queue us up?

MS. CRAWFORD: Sure. Go ahead and vote. We currently are looking for 14 votes until Emilio gets back on the phone. So once the polling responses reaches 14, then we'll know we have captured everything.

Is a number coming up?

Oh, it's fine. Once you've put your number in, yes, you should only have to do it once. We just need three more.

(Simultaneous speaking.)

MS. CRAWFORD: It is. No, no, at the top. It needs to warmup. So at the top, where it says responses, we have 12. But says polling open, so we need two more.

MS. MUNTHALI: You don't see it, typically you see it on the screen, you're not seeing it now.

MS. OGUNGBEMI: Can everyone point again? What are we at?

MS. CRAWFORD: Twelve. Yetunde, did you point --

MS. OGUNGBEMI: Yes. Oh, just for testing.

MS. CRAWFORD: For testing.

(Laughter.)

MS. MUNTHALI: So what are we at now?

MS. CRAWFORD: Still at 12. Thirteen. One more. Towards this laptop.

MS. MUNTHALI: Should be 14.

MS. CRAWFORD: Oh, maybe that's the issue.

MS. MUNTHALI: Oh.

MS. CRAWFORD: Okay, then we're good to go. Okay, we have our 14. Okay. Polling is closed.

And this is what it will look like. Thirteen voted yes, one voted no. And our percentage for each of that. So our clickers are working fine. We are good to go in terms of voting.

MS. OGUNGBEMI: Yes, so this is for real.

MS. CRAWFORD: This is for real.

MS. MUNTHALI: Okay, so the committee will be voting on performance gap for Measure 0032: Cervical Cancer Screening. This is a maintenance measure submitted by NCQA.

And the options are, 1 for high, 2 for moderate, 3 for low and, 4 for insufficient.

MS. OGUNGBEMI: Polling is open.

MS. MUNTHALI: Polling is open. Okay, we have 13 votes. That's the final. And the measure passes performance gap with one high vote, 11 moderate vote, one low vote. And the percentage, which will be shown very shortly, is 85 percent pass, with a moderate, and eight percent high and eight percent low. So we can proceed to reliability.

MEMBER HARRIS: So hopefully, Barry-Lewis Harris, hopefully reliability conversation will be very short because I think it seemed to be that the reliability was very high from what I read. I didn't see any evidence that it showed that the reliability was low.

MS. MUNTHALI: Are there any other comments? And as with evidence, reliability hasn't changed since the measure was last looked at. The committee can opt to not re-vote and just accept the previous decision from the last maintenance review. And I'm seeing nods saying yes. So we'll proceed to validity. Barry-Lewis?

MEMBER HARRIS: So the validity

testing, the actual threats to validity testing was related to the different measures. Whether the patient had a hysterectomy or no residual cervix or absence there.

The risk adjustment was actually none that I think I read about. And there were different numbers that were relating to meaningful difference of the commercial versus Medicaid families we discussed previously, a little bit before, I think.

And this particular evidence, I mean this particular measure, wanted to ask the developers whether or not they had any new information about maybe moving this one way or the other in improving the threats to the validity or if you feel that this is the best that we would have?

MS. ROTH: So I think we feel comfortable that there are no significant threats to validity for this measure. I also wanted to add that we, you had mentioned the exclusion for absence of cervix or total hysterectomy, and NCQA

does have a Policy Clarification System in which we take questions from health plans about the measure and respond to health plan questions.

And for this particular measure, we have not noted any questions that we've received that would make us question validity.

MS. MUNTHALI: Are there other questions? There was a, we were reviewing the committee's comments, and there was a question about the t-score.

And so there was some comments, also from staff, about the significance of that t-test to the measure score and how it was able to distinguish poor from high quality.

MS. ROTH: I can respond to that and provide a little more information. So this particular test we had determined the P value of an independent samples t-test and we compared commercial plans in the 20th percentile to the commercial plans at the 75th percentile. And what the P value of that shows, if it's less than .05, than the two groups performance is

significantly different from each other.

And so this is just another way of demonstrating that there is meaningful differences in performance across the health plans.

And so we did this for commercial plans. The P value was less than .05. And the same with the Medicaid plans, it was also less than .05.

MS. MUNTHALI: Okay, are there any other questions? Robyn, did you --

DR. NISHIMI: No. The staff recommendation was based on just face validity, but since NCQA has fought for the information on score level, it is eligible for high.

So you have three options for validity. Or because I think it's maintenance, you can also just, as with reliability, not vote again. It's really up to you.

MEMBER HARRIS: Did you -- I can't, I was looking for it, what was the actual recommendation for the last review?

DR. NISHIMI: We didn't have the same --

MEMBER HARRIS: Just for the record.

DR. NISHIMI: So it was a -- it was valid, it was strong. We didn't have the four-part voting system, but it was judged as valid. And so we would, we could go forward with that. Or you could vote, if you wanted to give it a four-parter.

CHAIR McINERNY: I think the sense of the committee is to vote. Thank you.

MS. MUNTHALI: Okay, so just wanted to, as Robyn mentioned, the highest rating this could receive, because it has face validity only, is --

DR. NISHIMI: No, they just explained the score.

MS. MUNTHALI: Oh, they did?

DR. NISHIMI: Yes, the t-test. So it's --

MS. MUNTHALI: So, Karen, now we have our methodologist coming here.

MS. JOHNSON: Well, I understand that that would actually show, as you mentioned, meaningful difference between plans. That is not usually what we're looking for in terms of score level validation. So I would still say that you'd be looking at the face of validity for this.

MS. MUNTHALI: So you will be looking at face validity. The highest it can receive is a moderate. This is a two.

So the clicker, the options have changed. So, 2 for moderate, 3 for low, 4 for insufficient. And so voting is open. Sheila, how many votes? One more vote?

MS. CRAWFORD: We have 13.

MS. MUNTHALI: Okay.

MS. CRAWFORD: That's what we're looking for now, right?

CHAIR McINERNY: There we go. Now we got it.

MS. MUNTHALI: Thirteen, okay. Okay, so this is not right. It didn't look right. So

let's do this again because what, for those on the phone --

Yes. So remember 2, 1 is not an option. One is not an option. Two is moderate, 3 is low, and 4 is insufficient.

So for those on the phone, all of the votes were low, we knew that wasn't right. Because they did meet, face validity.

(Simultaneous speaking.)

MS. MUNTHALI: You all hit 2?

(Simultaneous speaking.)

MS. MUNTHALI: So let's do a hand vote for this one.

MS. CRAWFORD: Yes.

MS. MUNTHALI: So if you can raise, if you're voting moderate can you raise your hand?

MS. CRAWFORD: Thirteen.

MS. MUNTHALI: Thirteen. It's unanimous. Moderate.

MS. CRAWFORD: Okay. Yes.

MS. MUNTHALI: Yes. So we'll look at that. Yes.

MS. CRAWFORD: Okay.

MS. MUNTHALI: So we'll move on to feasibility.

MEMBER HARRIS: So with feasibility, which is the extent to which it could be measured, the rating, I guess the preliminary rating, is related to moderate. And so could we have a little bit more information related to the concerns that were brought up by the staff? As to why this wouldn't be high versus moderate.

MS. MUNTHALI: Yes, I think part of the concerns was some of the data was out of E, or an electronic sources and they weren't defined fields. I think that was part of the feasibility concern that was raised. Robyn, did you want to speak to it?

DR. NISHIMI: Yes. It just said, so we quoted from the submission. It said sum data as opposed to all data. So that's why it went from high to, in the staff's opinion.

MEMBER HARRIS: And green, I'm not sure what that -- okay, great. More than three

people. What about the fact that someone had decoded, other than it actually being automatically put into the system?

DR. NISHIMI: That could be. I can just speak to what I would have indicated. That's often the case --

MEMBER HARRIS: Okay.

DR. NISHIMI: -- with administrative data. So that's taken as a given. But it was the sum data that was of higher concern.

MS. MUNTHALI: Any other concerns or questions for NCQA?

DR. NISHIMI: I just wanted to say, that's something that you're free, of course, to take into account though. The translation issue.

MS. MUNTHALI: So just a reminder, feasibility is not must pass. It is very important for us think about when we're thinking about maintenance measures or any new measures that come forward.

So the options, again, are 1 high, 2 moderate, 3 low and 4 insufficient. Voting is

open.

All right. So it looks like we're at 13. And so four voted high, nine voted moderate. And so this measure, 0032, passes on feasibility.

So we will now vote on usability and use. And, Barry-Lewis, any comments?

MEMBER HARRIS: Just a question for the committee, was related to whether or not it could be used, further goal of high quality and efficient care.

And the rating I see, preliminary was high, even though we did have a conversation related to only testing those commercial and Medicaid plans without actually looking at populations that are more higher risk. So for those populations that we actually are looking at, it would be a good high rating I would say.

DR. NISHIMI: Arjun raised the question about usability and use in the context of evidence and whether it got taught. But it's probably relevant to discuss here. Did you want to go over that, Arjun?

MEMBER VENKATESH: Sure. I mean, I think there's a tremendous amount of interest in the measure and a lot of people want to use it. That's the evidence that you see of all the accountability programs.

The measures used within it gets rated by the MAP, that this should be included in other programs.

And so that is what I feel like you have to balance with that table that basically shows the last three years' performance. And the mean doesn't change, the IQR doesn't change. And there's essentially been no change in performance.

Like it's really rare to see that little of change in a performance in a measure over three years. And so I'm going to probably rate this moderate.

And the reason being that, on one hand I see a measure that has a lot of conceptual interest in use and active use in programs, but on the flip side, I don't see really any evidence

that, since it was originally endorsed, that there's been any action on it.

Now if you could say, hey, there are, this measure was used during these quality collaboratives and there was improvement there. Like some evidence that somebody got better when this measure was used, I think that that would improve the usability rating. But that's just not there within the current presentation.

MS. MUNTHALI: Marcel?

MEMBER SALIVE: So my only concern on this is really, I think what Steve mentioned earlier, that there's sort of an over testing that's done like by a lot of people in practice. And I think that this measure doesn't really deal with it, and so it gives me that concern.

And I believe that that is done, you know, individually, by practitioners and anecdotally quite a lot. So this has no breaks on it. It just looks at the use.

MS. MUNTHALI: Any other comments? Okay, we will vote on usability and use for

measure 0032. High is 1, 2 for moderate, 3 for low and 4 for insufficient information.

Okay, 11 voted moderate and two voted low. So this measure passes, usability and use. And so we'll go to overall suitability for endorsement.

Okay, any last comments on anything that we've discussed before, before we vote on the overall suitability for endorsement? Okay, no comments.

One, yes, 2 no. Voting is open. Okay, two unanimous, 13 voted yes. Measure 0032 is recommended for endorsement.

Thank you, you guys did really well. It typically takes about 90 minutes for a committee to do the first measure, you were well under 90 minutes, so we will give you a 15-minute break. We can come back at 10:35.

CHAIR McINERNY: No, I just have one question for the measure developers. Would it be possible, going into the future, to correlate with HPV vaccine status?

Because I'm just wondering if some women who say, you know, I've had the three HPV vaccines and I prefer not to do the Pap smear, would that affect your results?

MS. BYRON: Yes, I think that's something that we can look into. It's interesting.

The HPV measure is for adolescents up to Age 13 though. So we might have to look at a different data search. I'm just thinking about how we would do that.

Because cervical cancer is a different age group. But we could still look to see if the measures are correlating. So that's interesting.

MEMBER HARRIS: I would just like to say, well, two things. One, to tag what Tom is saying.

The HPV vaccine and the measurement of HPV is supposed to, not the vaccine, but just measuring for HPV is supposed to lengthen the time. So you go from three to five.

So I think that's the point that was

being made is that if you're doing the HPV screening and then you're getting a negative result, then you say, hey, well there is more time available to not actually have to screen. And so that's what, I think, the question is. Are we looking at that, not necessarily the vaccine for the adolescent age?

Second, Elisa, first time out of the box, so there you go. That's the 90 minutes.

MS. MUNTHALI: You did great. You did great. Thank you. So we'll be back at 10:35.

(Whereupon, the above-entitled matter went off the record at 10:19 a.m. and resumed at 10:35 a.m.)

CHAIR McINERNY: Thanks everyone for reporting back promptly after the break. I believe, Michael Baer, are you on the line now?

OPERATOR: He is not.

CHAIR McINERNY: He's not. Okay. And how about Emilio, are you back? Are you here or are you on the line?

(Laughter.)

CHAIR McINERNY: Neither, okay. On route hopefully. Okay, so we're ready to go with the next measure please.

MS. MUNTHALI: And before we do that, Katie, I don't know if you were here when we went through and did introductions and disclosures --

MEMBER SELLERS: Yes.

MS. MUNTHALI: You did? Okay, great. So we just wanted to announce again that we will soon talk about the, we're talking about the childhood immunization measure right now, but Robyn will be recused from every measure after this, so she will not be participating in any discussion.

When the measure that she was involved in comes up, she will actually leave the room. And that is per our disclosure of interest policy that applies to, not just committee members, developers, but also to staff and consultants. Robyn?

DR. NISHIMI: I'm just going to actually leave as soon as immunization starts.

But I'll be back after mine comes up, which is the dialysis one, so.

MS. MUNTHALI: So our next measure to review is measure 0038, this is Childhood Immunization Status. This is also stewarded by NCQA. And we'll turn it over to NCQA, maybe Mary, can you introduce yourself?

DR. BARTON: Sure. Thanks very much. This is Mary Barton, I'm Vice President for performance measurement at NCQA. And Sepheen Byron is going to introduce the measure.

MS. BYRON: All right. So this is another health plan measure in HEDIS. It's a longstanding HEDIS measure that looks at the percentage of children who receive their recommended vaccinations by the age of 2, and it applies to commercial and Medicaid health plans.

It's also a measure that's widely used in programs such as the Medicaid Child Core Set. We use it in our health plan accreditation programs as well, and it's based on the guideline from the Advisory Committee on Immunization

Practices.

MS. MUNTHALI: Great, thank you. So Tom and Katie, you are the lead discussants. I'm not sure who would like to start.

CHAIR McINERNY: Katie, why don't you go ahead first, please, if you don't mind. Thank you.

MEMBER SELLERS: Okay, sure. So let's see, the developers gave the measure number and title. This is the percentage of children 2 years of age who have had four diphtheria, tetanus, pertussis, DTaP, three polio, one measles, mumps and rubella, three Hib, three Hep B, one chicken pox, four pneumococcal, one Hep A, two or three rotavirus and two influenza vaccines by their second birthday.

CHAIR McINERNY: And a partridge in a pear tree.

MEMBER SELLERS: Yes. And the measure calculates a rate for each vaccine as well as a combination rate. The numerator statement is children who receive the recommended vaccines by

their second birthday.

The denominator statement is children who turn 2 years of age during the measurement year, and the exclusions are children who have had a contraindication for specific vaccine from the denominator for all antigen rates and the combination rates.

So the denominator for all the rates must be the same. It's a process measure. The level of analysis is the health plan and the date it comes from administrative claims and electronic critical data.

And I think the developer said this, but it was originally endorsed in 2009 and most recently in 2012.

So moving on to the evidence. There's a systematic review, a ton of evidence here. There is some updated evidence. Previously they were relying on the 2011 ACIP recommendations and now it's the 2015 recommendations.

I guess the one weakness is there's no specific evidence cited for the measurement that

combines all ten immunizations. If you go through the evidence algorithm, it basically comes out to be moderate.

There's a question for the committee that I wasn't completely sure that I understood. So the question is the developer reports the guideline has been updated, but the 2015 update does not impact the measure and so is consistent with the specifications.

DR. NISHIMI: So then the question is the committee can as with the first measure forego voting again.

MEMBER SELLERS: Okay. Okay, so does the committee feel that it's okay to forego voting on the evidence based on the consistency of this measure over time.

MS. MUNTHALI: Okay, so we'll go to performance gap.

MEMBER SELLERS: Okay, so moving on to performance gaps, they provided the data on this for each of the ten vaccines and the combination rate of all ten. When they combine it they show

a mean of 47.57 for commercial plans and a mean of 36 for Medicaid.

Sorry, I was comparing the years. So an 11 point difference there. The same issue with stratifying by race, ethnicity or language. So then the question for the committee is, is there a gap in care that warrants a national performance measure.

MEMBER VENKATESH: I guess I'd just ask a question of the developers here to understand performance on this measure. So the new part of this is the composite part. Is that true?

MS. MUNTHALI: No.

MEMBER VENKATESH: It's always been there?

DR. NISHIMI: No. The composite was previously there. They actually had many more composites.

MEMBER VENKATESH: Okay, then I'll hold my question for validity.

MS. MUNTHALI: Matt?

MEMBER STIEFEL: Can we just make our recommendation about race, ethnicity and language and maybe SES to apply generally to all of these for which it's relevant?

MS. MUNTHALI: It's noted. Other questions, recommendations?

Okay, even though this is a maintenance measure we still have to vote on performance gap. So 1 high, 2 moderate, 3 low, 4 insufficient. Voting is open.

Okay, so 11 voted for high, two moderate, so this measure passes on performance gap. And so because this is a composite we need to assess the construct of how the ten components of this measure hang together.

So there's a vote on that and discussion first. So Katie and Tom.

MEMBER SELLERS: So the discussion questions here are is the quality construct logical combining the ten recommended vaccines, and I would say yes unless anyone has any comment about that.

Is inferring the individual ACIP recommendations apply as the rationale for all ten composite appropriate? Again this has been addressed by the committee previously, I believe. I don't know if there are any comments about that.

DR. NISHIMI: Actually, the previous forms didn't break out the composite this way, so that's why we do need the committee this time. Even though they had been endorsed --

MEMBER SELLERS: I see. This is a change in the NQF procedure even though the measure has not changed.

DR. NISHIMI: Right.

MEMBER SELLERS: Got it. Okay, so is there any discussion about the appropriateness of combining all ten vaccines?

MEMBER HILL: I think there were some discussion previously about it being an all or nothing, being a disincentive for getting, you know, people who were accomplishing 75, 80, 90 percent were feeling like they weren't getting

credit for what they had done.

I don't know if anybody else has heard that feedback. Oh, sorry. Steve.

MEMBER TEUTSCH: To circle around that same line, the actual utilization of all these individual vaccines is really, really high and it does look, you know, like the gap is really big, when in fact it's actually relatively modest. And for many of these diseases the rates now of disease are extremely low fortunately. This is obviously one of the great successes.

So I'm not against this measure in any way, but it does suggest that you may lose some of the focus on the areas that actually do need improvement, because it does suggest a bigger problem than probably exists.

MEMBER HILL: I would agree and that's consistent with the feedback I've gotten. And I think, you know, with a successful measure we have to be willing to kind of drill down and see what within that composite now constitutes the primary gap.

MEMBER TEUTSCH: Can I say one more thing sort of along those lines? I guess one could have a measure saying how many kids meet 90 percent or more, or 80 percent of more of all these vaccines. Perfection is pretty hard to achieve.

And while I'm not suggesting any one of these should be removed from the list, there are some that, you know, are tough. And I also worry, frankly, about the immunization measures that the biggest problem is people who opt out right now, and it's become a bigger problem in oftentimes in the communities you think should know better.

But that's not really addressed in here, which is what's really, in addition to getting those who aren't getting them for whatever clinical reason you have this preference stuff which at least in some states has gotten pretty much out of hand, or some communities.

MS. MUNTHALI: Arjun.

MEMBER VENKATESH: So I'm not a

clinician that ever immunizes kids and so I don't know that part of it, but I guess the question I ask thinking about it from that perspective is when we have a composite that says do all ten, what that conceptually says is that these are all ten equivalently important.

And so I understand the concern that it's going to detract attention. And so if you look through the ten measures, the components and the performance rates, the 25th percentile exceeds 80 percent for, I think, seven of these, the vast majority.

The ones where it doesn't exceed that number, the lowest is influenza and then to a lesser degree rotavirus and pneumonia. And so what that makes me think is when I see the composite scores that are quite low that basically, probably means that flu immunization is what is driving the composite to be low.

And so the question I would ask you all to think about, and I would like to hear what people who actually do this clinically say is, is

would getting a low score, seeing a 38 percent, 40 percent in this measure impede the credibility of this measure when you know that your performance is high for everything except flu and what are the implications of that?

And the second, I guess, question for the developer that's related to this is have you done analyses that compare in developing the composite that look at the relationships between performances -- I'm sure you have -- performances on several combinations with other combinations, one measure with the other?

And so is composite performance variation largely explained just by flu vaccination or is it actually a different world where different plans are underperforming at truly at different measures and the composite is driven by differences in performance?

MEMBER HILL: I can tell you from talking with hundreds of physicians that they have a real conceptual problem with this measure because it's a composite measure and they do not

measure their performance using this.

And so it fails to engage the frontline providers when you're trying to drive quality improvement for these very reasons that have been mentioned. This is mainly in Texas and Florida.

MS. MUNTHALI: Matt.

MEMBER STIEFEL: I just wanted to follow up on Steve's comment and ask the developers if, and maybe it's in here but I didn't see it. Do we have the percentage of people who opt out?

The reason is that potentially becomes a different measure, actually a health population, health and well-being measure about public health efforts to educate people about the importance of this immunization.

MS. BYRON: So I've heard a couple different things that sound -- but see if I can get most of them. So just to be clear, NCQA actually reports this measure out for each individual vaccine rate, and then we also have

different combinations.

And it's actually not just combo 10, we have combinations of all the different vaccines as you go and we report them all out nationally.

So, you know, if a health plan wanted to benchmark against combo 5 which is, you know, the first five, then they could do that. I think the value in a composite that I'm hearing from everyone is that you can use it in different ways depending on the program.

And in fact we do use the combinations in different ways according to the program. For the longest time combination 2 was the one that was in our health plan accreditation program, but then we realized that we needed to raise the bar and go beyond that. And so then we did push for combination 10 to be used at the health plan level.

In order to understand which combinations to use, we did do some analyses on how they hang together so that is a good point.

You need to understand which combinations are going to be the most useful.

But in order to allow the flexibility and the comparisons for health plans we do report out each individual vaccine rate and then we do report out the combination.

So if a health plan wanted to say what is pushing our rate really low, you could look and you could look across health plans to see how others are doing on rotavirus, you know, which does tend to be one that is a little lower than something like flu.

So the measure allows for that flexibility. It's really about whoever's implementing the program to figure out which combination that they think is best to be using for accountability or quality improvement or what have you.

So this is a health plan measure and across HEDIS we actually do not allow a health plan to say, well, someone refused and so I get an out on the measure.

You know, for this measure and many others in HEDIS we feel that it is really the responsibility of the health plan to get this done, and so if someone refused you don't get an exception to it. And we felt that at the health plan level that was appropriate.

We do report out regionally as well so that you can look, you know, for those regions where you are seeing a lot of refusals crop up you can look geographically to see how things might be impacted in places like the Pacific Northwest. So, you know, because we do hear about those pockets.

But, you know, we report out the measure with the hopes that people will continue to encourage everybody to be getting their vaccinations. So that's the refusal and there may be others.

CHAIR McINERNY: I argue strongly to use the combination of all ten vaccines, because in this day and age of vaccine hesitancy and refusal if we start to say to plans and it then

will filter down to physicians, well, this wasn't so important, we're going to have real problems.

Because we know already that if we drop below 90 percent, and this has happened several times over the past few years, the most recent and famous of which was the measles in Walt Disney -- the one in California whichever one that is, and -- Disneyland, I guess -- and, you know, and we have problems with infants getting pertussis, infants who cannot be immunized before age 6 months getting pertussis from older children and adolescents who are inadequately immunized and so forth.

And so I think it's very important that the plans be, continue to be incentivize to measure all ten vaccine rates.

MS. MUNTHALI: Steve, did you have a comment?

MEMBER TEUTSCH: I was just going to make the point that not all these vaccines are equally effective either, and the strength of evidence for some of them such as flu in younger

kids is not all that great.

And I'm sure that leads to some of the differences, so actually I felt better after you said that you can look at these in different ways.

But it is a problem when you sort of think of them all the same, because there are so many of these critical ones as Tom was saying where you've got to get them up to high rates of immunization if you're going to get herd immunity.

CHAIR McINERNY: Steve, that's a point well taken. And as probably most of you know, recently the ACIP and the AAP agrees with them, is that intranasal influenza vaccine is not effective and should not be used.

I think this is something we need to keep in mind for our flu vaccine measures that are coming up. I'm not sure whether an intranasal flu vaccine is considered appropriate. It should not be from now on. The only vaccine that should be considered appropriate should be

the injectable.

MS. MUNTHALI: Arjun.

MEMBER VENKATESH: So I am totally sold on the conceptual framework that if we promote a full 10 combination that if you get one vaccination you're more likely to get another vaccination would be a reason enough to support the composite.

My fear is this, and I think if the developer could share just the correlation between each individual measure in the composite that would be reassuring enough. Because if the correlation between just the flu vaccine and the composite is exceedingly high, north of .8, let's say, then it tells me the composite's really not measuring anything new or different, it's just measuring your flu vaccine rate.

Then you have to make a decision, which is this kind of policy or sort of a more quality decision is do you think that encouraging flu vaccination encourages the other vaccines, or do you think that encouraging flu vaccination may

create an ala carte mentality?

Oh well, okay, there's so many of these different shots let's do the flu shot and maybe not another one that has more evidence base to it.

And so I don't know that world of how people react well, but I do think it's if, I think you've got to, you have to know if the composite's actually measuring something different than flu rates or just flu rates.

MEMBER HILL: I think that observation on flu is important because basically you're starting every year with a hundred percent gap on flu. So it's a very different type of immunization than the others, right.

MS. MUNTHALI: NCQA, perhaps you'd like to address Arjun's point about the feasibility of having individual rates for the individual composites.

DR. BARTON: I guess I would say we don't have that data at our fingertips at this moment. I would be glad to go back and try to,

and run that but I'm not sure actually that a correlation across all plans who report means that there's not a part of the health care system that doesn't have a different experience.

I'm a little bit struggling with your logical model here. I can appreciate that in general if there was very high correlation or even a hundred percent correlation between influenza and the composite, or the 200 health plans or 600 health plans or however, you know, many, many health plans that NCQA assesses that that would be suggestive of a trend or a predominance.

But it's hard to imagine that it would actually come back as a one to one, and even if it did does that mean that there's not any part of the health care system that doesn't benefit from having an NQF endorsed composite measure to drive quality and accountability in that sector of the health care system?

MEMBER VENKATESH: I guess I would just say that if the correlation, if there is one

like you would suppose, then they're not measuring different things they're measuring the same thing.

And so we need to have some degree of transparency about that and just, you know, people should be aware that okay, my composite 10 score is 30 percent but that's because my flu rate is 30 percent.

And so I guess partially that's addressed by the fact that you report out individual measures and people get all that individual information.

The flip side is on the use side. If you found people using the composite 10 and not using sub-combinations for use applications, then you could imagine a world where based on whatever the use of the measure is -- accreditation, certification, payment, whatever it is -- if they only used a composite 10 they may be thinking they're actually getting a full 10 composite, but really they're getting the same information as they get from just the flu measure.

And so that's, part of this is the intersection between the use of the measure and how this is kind of, you know, currently presented.

DR. BARTON: And we can only speak to how we use the measure and by offering measures to be NQF endorsed and thereby offering them on the Quality Positioning System. They're available to people, but it is I would say impossible for us to police every use in every locus of a measure that we've developed.

MS. MUNTHALI: I just wanted to piggy back on what Mary was saying. We should be looking at the construct of the measure and the scientific merits of the measure.

While our process we say is use agnostic, we do understand that issues -- you live in the world. We all live in the world and so you will consider, you know, the unintended consequences, but we'd like you to focus on the composite right now on 1c looking at all of the issue you brought up.

Some of them will be also addressed in 2d when we talk about the rationale for our hanging these composites, these components together. But just, Mary, that was a good reminder for everyone.

MEMBER HILL: So does this decision that we're making assume or include, the intent is that it includes, there's evidence that supports the logical clustering of these is what we're saying, right?

MS. JOHNSON: Yes. It's kind of tricky the way we have composite measures in our criteria because we're actually splitting it up into two pieces. The first piece under 1c is just the rationale for doing it.

So in other words, why did you put these ten together? Why wasn't it 11, why wasn't it 9? That kind of thing, right. And Arjun, you're very correct. This is an all or none measure, so the waiting is equal.

So we just want to make sure that you guys understand NCQA's rationale for why they did

what they did and that's what you're thinking about in 1c.

When we get to 2d we're going to go a little further and we're going to hit Arjun's other question which is, is the composite, the 10, being driven by perhaps one or two?

The question there is if it's being driven by one or two is there a reason to have the composite, why not just have the one or two? That's the question that you'll be looking at under 2d and you'll be looking at the actual data that the developers provided.

I don't think, when we get there I don't think you've done those correlations statistics. They would be a little bit of interest, but you've given the individual performance rates for the different components and you can infer from that what's going on.

So did that clear up anything or did that add additional confusion to the committee?

So it might be useful just to have Mary or -- I'm sorry. Sepheen, just say one more

time, you chose to do an all or none with these ten. Just the elevator speech, why did you do that?

MS. BYRON: Well, and I do want to emphasize that we have all of them individually and that we report them out nationally by health plan and then we actually do all the different combinations, not just the ten.

On the form we submit a combination 10 because it, you know, I think when it comes to vaccines there is a thinking that these are the vaccines you should be getting by the age of 2, each is important for different reasons.

We have seen rates for different ones decrease in different times, you know, as Tom pointed out MMR one year, pertussis another, and so we do feel that it's important for all of the rates to be, or all of the different vaccines to be reported out.

And then the combination 10 we feel is a helpful way to look to see at health plan level is everyone getting their needed vaccines by age

2. But if you wanted to you could look at combo 5 or 6 or 7 or et cetera.

MS. MUNTHALI: Katie.

MEMBER SELLERS: I just had a question for the staff here. I was a little confused by where it said the rating is insufficient and it says change here. Is there some significance to that?

DR. NISHIMI: No, the change here was notes from something else. I noticed that too. It just didn't get cleaned up.

MEMBER SELLERS: Okay.

MS. MUNTHALI: Any other questions, concerns, questions for NCQA or NQF?

Okay. I think Marcel left, so Ann is 12. Okay. We still have four.

So this is 1c for a composite and it's a composite explicitly articulated and logical, the quality construct including components, 1c2 rationale for distinction additive value, 1c3 aggregation and weighting.

So your options are 1 high, 2

moderate, 3 low and 4 insufficient. And voting is open.

(Voting.)

MS. MUNTHALI: So for measure 0038: Childhood Immunization Status by NCQA, three voted high, four voted moderate, four voted low and one voted insufficient. So this takes us into the grey zone.

This means that consensus was not reached but we continue with voting and we'll resolve this. It's a major criterion and we'll resolve it on the post-comment call. So let's move on to the liability.

Katie and Tom, would you like to lead discussion on testing?

CHAIR McINERNY: Katie, go ahead. Thank you.

MEMBER SELLERS: Okay. Okay, for reliability -- sorry, just catching myself up here. Okay.

So the questions are, are all the data elements clearly defined and are all appropriate

codes included? It seems to me a very clear definition for this combination.

And then the next question is, is it likely this measure can be consistently implemented? I think we've seen that it has been. On the testing they use the beta-binomial method to assess signal to noise.

A reliability score of 1 would be perfect, zero would be completely random. The score on this from the 2012 committee was reliability statistics ranging from 0.84 to 0.98 depending on the vaccine.

So it sounds to me like that's not actually the composite measure that we're looking at. Is that correct, these data?

DR. BARTON: I'm sorry. You mentioned 2012 so I'm trying to find what you're quoting because we quote the 2014 data set. Data, is that what --

DR. NISHIMI: Yes, we pulled the reliability from the previous submission. But the --

MEMBER SELLERS: Here's the 2014, sorry. 2014 reliability statistics for all ten vaccines was 0.98 for commercial plans and 0.96 for Medicaid. So that's quite high. And when you use the reliability algorithm that also brings you to a rating of high.

So then the questions for the committee are do the results demonstrate sufficient reliability so that differences in performance can be identified?

The previous committee concluded reliability was high with reliability statistics of 0.84 to .98. The updated testing reveals reliability statistics of 0.89 to 0.98. Does the committee agree there's no need for repeat discussion and voting on reliability?

MS. MUNTHALI: We're seeing nods, yes. So we'll move on to validity.

MEMBER SELLERS: Okay. So the first question is are the measure specifications consistent with the evidence, and the recommendation here is yes.

If we look at the validity testing, let's see. So the last time this was reviewed in 2012, they provided face validity testing. This time it's still face validity only but the developer submitted a t-test and it's a little bit unclear how the t-test relates to validity.

So I would like to ask the developers to clarify that relationship there.

MS. BYRON: Yes, sorry. That was our error. We meant to put that under the test of meaningful differences section which is 2b5. So sorry about that confusion.

MEMBER SELLERS: Okay. So I think what we're looking at here is the same validity information that was provided in 2012 which is face validity only.

MS. MUNTHALI: And again, the committee can opt to just accept their previous decision from 2012. Is that a yes? Nodding, yes. So we'll move on to 2d.

MEMBER SELLERS: 2d?

DR. NISHIMI: Yes. There was an error

in the construction of the PA, so Karen is going to have to walk through. It really goes to Arjun's question about what's driving the lower rates in the all 10 and whether that undermines the construction or whether it's fine.

Karen, can you walk through the composite issue?

MS. JOHNSON: Sure. So it might help if we could pull up the 2d criterion so you can see how you're voting on 2d on your voting slides. Yes. This is our mistake and apologies for this. We forgot subcriterion 2d on your PA, so sorry about that.

But again just to describe what 2d is about, the question is -- you can see it on your screen on the voting screen. Four composite measures we want to know is there empirical analyses to support the composite construction and demonstrate that the measures fit the quality construct, add value, parsimony to the extent possible, aggregation and weighting fit the quality construct and simplicity to the extent

possible.

These things get really complicated when you have really complicated composite measures. In reality, this is a pretty simple all-or-none composite measure, right.

So basically the quality construct is that they've decided that these ten components go into the one composite, and we just want to know is there any empirical analysis that really supports that?

So in other words that the composite itself is telling you something different than what you would know from the individual measures, okay. Again it really is Arjun's question all over again.

Now they -- I don't think, and this is something that is also a little bit confusing because this measure is older than our most recent composite guidance, okay. We did this composite guidance, I think, in 2013, so quite a bit after you guys had done this before.

And in our effort to try to make the

submission process a little less onerous for developers we did not insist that they update to our newest forms, which means that they also didn't fill out 2d for you, okay.

But what they have is data that a lot of people do provide for all-or-none composites. Often for an all-or-none composite, if they're using our current form for 2d they will just show you the rates for the individual components. You have that under 1b, under gap, right, because they told you what the performance rates were for all of those individual components.

So you can look at that and I think really the question is, is all of those necessary to the composite, is anything kind of extraneous to the composite, is one of two things pretty much driving the composite?

You may be able to tell that from the performance data that were provided under 1b. It would be interesting to know if NCQA might be able to do that correlation analysis that Arjun's requested. That might tell you something a

little bit different than the performance rate.

You can decide if you need to see that before you could rate or not. But I shouldn't just assume that NCQA would be able to do that analysis either. So that would, I think, be a question for you guys.

MS. JOHNSON: Within, I think within probably a two-month period, something like that.

DR. BARTON: We'd be glad to do those analyses, and as you know we're glad to fill out the forms that you ask us to fill out. And so we could have done it in advance if we'd known that this was going to be part of it. Although in all honesty, as Sopheen has pointed out -- I'm sorry -- we have a number of ways that we present this data.

Back to plans, each individual rate, a number of composites. We could withdraw composite 10 from being considered and just have the individual rates in our proposed measure, then it's not in NQF's Quality Positioning System.

I see Tom shaking his head, but I would say that it was not our suggestion that this be considered as a composite. It was NQF's suggestion.

MS. JOHNSON: And just to make sure everybody understands, we did an expert group come-together a few years ago and they all agreed these all or none measures, which are kind of a different animal than the individual measures, we also want to think about them as composite measures.

And it is a fair question, you know, are the things in there useful. I think in terms of, you know, pulling composite out of here that would certainly be an option, but if you feel strongly that that composite, that 10-point composite is really offering something, you know, then you probably wouldn't want to do that.

And I would also say, you know, that extra data may not be needed by the committee. They may be able to look at your performance rates and not even ask you to do anything.

So, Arjun.

MEMBER VENKATESH: Karen, is it possible, do they have, do you all have face validity testing for the all-10 composite? Because if you have that could we use that to say, A) that's good enough to say that the composite's valid, because people say you should measure all 10 together as an all or none, and then ask that they bring data later. Is that an option?

MS. JOHNSON: It could be a compromise. I think I would prefer that you look at those performance rates and just make sure that you feel like something from those performance rates are helping.

We actually see 2d as something a little different than just validity testing. It's sort of validity but it's something separate so it's its own subcriterion. That would certainly bolster, you know, I could see some people saying that would bolster their comfort.

MS. MUNTHALI: Steve.

MEMBER TEUTSCH: Yes. Along the same lines, because so for full disclosure I was on the U.S. Preventive Services Task Force for a long time and obviously we have a lot of problems with being incomplete on lots of different preventive services.

But for the interest of parsimony as we get to some broader measures that can be used at a population level, having things like up to date on your clinical preventive services so it differs in exactly what those are, what based on your age and your gender and other things, but having that as a measure of how well you're doing is, you know, at some level those are more informative at population levels than all the detail that you get mucked in if you've got to go and deal with this in a real world in a plan or something where you want to see exactly where your deficits are. It's a different animal.

And so I would say that in general moving towards some of these larger scale, broader composite measures offer a lot at a

population level.

I know this is sort of in the clinical stuff, but if you're really trying to deal with stuff at a population level this is great, because then you've got to drill down and you start answering the questions that Arjun's asking.

All right, which one is it that I've got a problem with? Who, which are the people? Which is the geographic areas? So I find these things actually pretty helpful to deal with things that are at a more population oriented level because it's a place to start.

If I'm doing great on this measure, good. Then I'm going to go onto something else where I've got a problem. And if I've got to keep getting buried down in all the detail all the time, it really bogs down a lot of the processes and makes some of the communications on priorities and other things more challenging.

MS. MUNTHALI: Patricia.

MEMBER McKANE: Yes. I think my point

is almost the same and I hope it's a little bit different, but as an epidemiologist when I see composite I just kind of, I cringe because I think, you know, it's as Arjun said the weighting and some are more important than the others.

But I really do see a value with this as Steve said, you know, to measure you're measuring health plans. This isn't measuring the population, but yet it's kind of giving us an indicator or a signal of what's going on in the population.

And I think that we might just have to be clear and it behooves us to know it's a composite. That there are other factors that are going into it and that we do need, as you said you report out on all of it, you know, and give that information. Because there is value in the individual, but also looking at it as a group.

And I think just from my point of view is making sure, and I'm not the expert here so I'm really trusting that these immunizations, these ten vaccines are the ones that are critical

and they are all important to be completed by age 2.

MS. MUNTHALI: Marcel.

MEMBER SALIVE: Yes. So just, I guess, one small piece to add on is that I think we've discussed this in the past, I think, at this committee about harmonizing with the population measures.

And so I think there is some link here to Healthy People 2020, where it's not exactly the same but it's got some composites in there which I think are useful and it does also look at the individual ones.

So, but, you know, to Steve's point I think, you know, the plans contribute to that. Everyone contributes to that and so we have to, you know, having that harmony a little bit or some way of looking at it is helpful for all the local efforts.

MS. MUNTHALI: Arjun.

MEMBER VENKATESH: I'm trying to figure out some way to help you guys through this

and like where we can like, because I think the challenge you're all having is that we've got ten individual measures but there's a bunch of combinations, right.

So we're doing like one vote for some exceedingly number high of combination of measures. And so what I'm stuck with, I guess, is that the question we have in front of us is about what's the incremental values, their unique value of a composite versus the individual or versus the other measures?

I'm looking at two composites, what you guys call combination 7 and combination 10. The only difference in those two measures is flu.

And so I'm trying to figure out how to, what I think is conceptually makes a ton of sense to do an all-or-none, what conceptually makes a ton of sense is it seems to be consistent with a variety of other recommendations.

I don't know the guidelines and the various possible recommendations as well, but to me it's just then the combination 7 versus

combination 10. The only difference in those two composites is flu or no flu.

And so either you could say, hey, conceptually we pick and say endorse 10, which is the one that's in front of us, because it sort of makes sense and there's some face validity to having it together, or if you really hold us to the question of is there data to suggest that combination 10 is parsimonious and provides incremental value of a composite, I can't answer that question with any high certainty.

So I'd probably be left with a low to moderate, but as far a combination 7 then, just guessing basing it individual measure scores I'm more likely to be in a spot where the composites tell me more than I get from individual.

DR. NISHIMI: Arjun, I just need to clarify something for you and the committee. NCQA's previous submission included all those other combinations.

They clarified that for this maintenance mission they're only seeking, if you

look at the first page not all the other data, they are now only seeking endorsement of the ten individuals and the one all-10 now.

That's all that this particular, the reason you see the other data is because it's the old form. But this committee's deliberations are now focused just on the individual ten and the all-10.

MEMBER HILL: And if I can just add on to where you're coming from, Arjun, I've sat in on the health plan negotiations where the NCQA-endorsed measures be turned into insignificant influences on your fee schedule and how things are going, you know, at the patient level.

And so if we endorse 10, what I suspect is the next negotiation I sit in on, 10 will be the only thing on the table, not 7, not 5, not 1, no matter where you are.

And so that's going to affect a one to three-year contract. I mean, this is some of the practical things that happen when we endorse a composite measure.

MS. MUNTHALI: Other questions, comments or concerns? Okay.

OPERATOR: If you would like to make a comment, please press star 1.

MS. MUNTHALI: Only for the committee. We're not at a public comment yet.

Okay, so we'll be voting for measure 0038. This is the Childhood Immunization Measure by NCQA, 2d. This is, you know, making sure there's an empirical analysis that's supported by the composite construct and it's demonstrated

And the options are 1 high, 2 moderate, 3 low and 4 insufficient, and voting is open.

(Voting.)

MS. MUNTHALI: All right, for measure 0038, three voted high, seven voted moderate, three voted no and zero voted insufficient. So this passes, 2d, and so we'll go on to feasibility.

Katie and Tom.

CHAIR McINERNY: Katie, you mind going

again? Thanks very much. You're doing a great job.

MEMBER SELLERS: Sure. Thank you. So we're moving on to feasibility. This is already in use, widely used. There are no issues identified regarding feasibility.

MS. MUNTHALI: Comments from the rest of the committee?

MEMBER HARRIS: I just wonder why it was coded as moderate versus high.

DR. NISHIMI: For the same reason as the last one because it says some data.

MEMBER HARRIS: Okay.

DR. NISHIMI: It's a judgment call. The committee can choose to go high.

MS. MUNTHALI: Okay. It looks like we're ready for a vote. So for feasibility, measure 0038, 1 high, 2 moderate, 3 low, 4 insufficient. Voting is open.

(Voting.)

MS. MUNTHALI: Okay. So four voted high, nine voted moderate. So for 0038 the

committee has passed on feasibility, so we move on to usability and use.

MEMBER SELLERS: For usability and use it is reported that this is currently used, currently publicly reported, used in an accountability program. Well, they list six accountability programs including NCQA health plan rating.

For improvement results they do note that during the last five years performance has improved across commercial plans. The proportion of children documented as having received all ten vaccines moved from less than a fourth to about half. For Medicaid plans it went from about 15 percent to a little over a third. The 2014 rates are still showing room for improvement, 47.6 percent for commercial plans, 36.1 percent for Medicaid plans.

They note that receipt of some individual vaccines is high, while others remain low. Large differences between the lower and higher performing plans exist. The range among

commercial plans -- sorry.

The average rate among commercial plans was 28.4 percent in the 10th percentile, and up to 63.2 percent among those in the 90th percentile. For Medicaid that range was 23.4 percent to 49.6 percent.

The developer's not found any unintended consequences during testing or since implementation. They don't report any potential harms.

They do note that -- the 2012 committee noted that in times of vaccine shortages, adjustments or explanation may need to be made for the rates.

And the question for the committee is: can the performance results for this measure be used to further the goal of high quality efficient health care? And the staff preliminary rating for usability and use was high.

Questions or comments?

MS. MUNTHALI: It looks like we're ready for a vote. So one, high; two, moderate;

three, low; four, insufficient information. Voting is open.

(Voting.)

MS. OGUNGBEMI: Michael, if you could submit your vote via the chat, please.

MEMBER BAER: It should be there now.

MS. OGUNGBEMI: Received.

MS. MUNTHALI: So, we have 13. Yes, 13. So 12 voted high; 1 voted moderate.

So as you remember, we did not reach consensus on 1(c). It's a major criterion, so we cannot proceed to an overall vote. We will resolve that issue during the post-comment call, and we'll further discussion of that criterion, re-vote on 1(c), and then do an overall vote.

So that post-comment call is in about two months or so, two and a half months. December 6th, okay. So thank you. So we are -- oops, Matt.

CHAIR McINERNY: Just one quick question for NCQA. Have you at all correlated your results with the state registries? Do you

find them helpful? Do you think there would be any value in doing that?

MS. BYRON: That's a good question. And for this particular measure, actually, I'm not sure we have, but recently for the HPV vaccine measure because that is one that we had been updating the most recently, we did look at state registries.

And we did find that the results in the field testing among health plans compared to the registries were similar, so you did see the same vaccines being higher in one and higher in another and, you know, vice versa.

The state registries vary widely across the U.S., you know, some are very good; some are not and we are hoping to get to a point where the registries are all giving really good information.

I think the primary issue is how you handle certain things like accounting for the denominator. So a lot of state registries have people in them, but when they move out of the

state they may not be removed.

So dealing with issues like that I think we still have to allow the registries to catch up, but we do allow plans to use registry data to report the measure, and they're all audited, so --

MEMBER STIEFEL: Just a question for NCQA for future consideration, and that's whether you've considered a weighted average composite, which would maintain the value of parsimony but increase the validity of the measure because of the -- and weighting may be based on impact on population mortality and morbidity.

MS. BYRON: Thanks for the suggestion. You know it's interesting because with the vaccines, you know, and the combinations, we've looked at other measures and composites such as for well care with children, you know, looking at BMI and then counseling and then risk assessment.

And definitely it's more complex in thinking about whether everything should be weighted the same or whether it should be all or

nothing or should it be around, you know, opportunity to provide the service.

For vaccines, I think we felt it was a little bit more straightforward, but I think we can look into that.

DR. BARTON: And you know, we've certainly been interested in the panel's feedback. I guess what's hard to understand is that the ACIP is seen, coming from the CDC, to be the foremost authority on what age-appropriate vaccinations are.

And so as we are consumers of guidelines and recommendations from august bodies such as ACIP and USPSTF, it's in -- while there are potentially interesting scientific byways to go down, it seems like the further we get away from those authoritative bodies the harder it will be to come before the next NQF panel.

MEMBER TEUTSCH: So I worry actually about ACIP, just like I do with groups that are constituted by other professional organizations that actually are very committed to whatever the

technologies are.

I don't think it differs all that much. These are people who are very committed to vaccines as the solution. And there are a lot of people who are there who are heavily invested.

And so I don't think it differs radically from what you see from other professional groups that have similar kind of vested interest. They tend to be pretty aggressive about approving things in my estimation.

I know they are the authoritative body. I can't tell you where else to go. But I would think that if you actually had a body that was constituted somewhat differently, that didn't have as its mission basically just the control of diseases, but looked at all the data in a different type of critical way, you would end up with different recommendations.

As we're going to talk about in flu, some of the data are not all that strong in some of these age groups. So there is a willingness

to extrapolate, and to groups that -- others who are looking at a different level of critical, they might come up with different things.

The Brits have different, who are very evidence-oriented, have different recommendations, for example.

So I don't think, I don't have any wisdom to shed on what you should do other than use them, but I do think we have to realize where all these recommendations do come from that we, you know, we ensconce sort of in these metrics.

MS. MUNTHALI: Thank you. So we will move on to the next measure which is 0039. This is the flu vaccines for adults ages 18 and older. This is also stewarded and developed by NCQA.

Before we begin, there are a couple of things we need to take care of. First, Michael Baer is on the line. He's our committee member. He missed our intro and disclose of interest. So Michael, I'll ask you to introduce yourself and let us know if there's anything you'd like to disclose in relation to this.

MEMBER BAER: Sure, thank you. I'm from AmeriHealth Caritas Pennsylvania. It is a Medicaid managed care company in Pennsylvania, and I have no disclosures.

MS. MUNTHALI: Thank you very much. And on the heels of Steve's last comment, we did have a suggestion to make the next round of influenza vaccine measures more efficient.

Staff has determined that the evidence base for all of them, they're based on the ACIP guidelines, and we pretty much have given it the same evidence rating across all of the measures.

I think there are eight influenza measures. And what we would like to recommend, if you'd like to take it up, is that we could vote. You could either decide to vote on evidence for all of them knowing that same basis for evidence, or you can decide to forego conversation on evidence for all.

Well, there are seven maintenance measures, one e-Measure, but it is the evidence basis is based on a current maintenance measure,

claims based measure, or we can have discussion on evidence for 0039, and that vote can carry over to the remaining influenza vaccine measures.

So we can either forego that for those of you who've reviewed the evidence for the influenza vaccine measures, if you'd like to comment on that. So Marcel.

MEMBER SALIVE: Well, I'd like to have a small discussion on evidence. You know, I don't think I want to discuss each one individually.

MS. MUNTHALI: Okay. So let's start off with the measure before us. So Mary and Jenna.

MS. WILLIAMS-BADER: Hi, my name is Jenna Williams-Bader, and today I'll be talking about our flu vaccinations for adults 18 and older.

So this measure assesses the percentage of adults 18 years of age and older who self-report receiving an influenza vaccine since July 1st through when CAHPS is

administered, because this is collected using a CAHPS question.

There are two age stratifications, 18 to 64, and 65 and older. It is a health plan level measure like the two you've already discussed, and it's reported for commercial and Medicaid for the 18 to 64 age population, and then Medicare for the 65 and older.

This is a very longstanding measure, and at NCQA we do still see room for improvement with Medicaid performance about 40 percent, commercial about 49 percent, and Medicare about 73 percent.

And this measure is used in a number of programs including NCQA's health plan ranking and our accreditation, the Medicaid Adult Core Set, and Medicare Stars.

MS. MUNTHALI: Great, thank you. So I think the lead discussants on this one are Catherine Hill and Michael Baer, so I'm not sure who wants to start off.

MEMBER HILL: I'd be happy to. Are we

going to skip the evidence on this?

MS. MUNTHALI: So Marcel had put a motion out that we at least talk about it a little bit, yes.

MEMBER BAER: I just want to say thank you, Catherine. This is Mike Baer.

MEMBER HILL: No problem. All right, so on the evidence, we see that it does have a systematic review that the quality, quantity, and consistency of the evidence provided was good. That it does not have a graded evidence.

The developer has updated the evidence for this measure, adding the 2015-2016 ACIP recommendations, and the measure of course as you've heard remains aligned with those recommendations.

The question for the committee was the developer has reported that this guideline has been updated, but the 2015 update, and that 2015 update does not really impact the measure, is deemed consistent with the specifications.

And as the committee, we are tasked

with trying to decide whether we agree, whether there's no need for repeat discussion and voting on the evidence. And I hear from my colleagues that we do want to discuss the evidence. It's got a preliminary rating of high.

MS. MUNTHALI: Michael, anything to add?

MEMBER BAER: I have nothing to add, thank you.

MS. MUNTHALI: Okay, Marcel.

MEMBER SALIVE: So I reviewed it for a different measure, too, but I wanted to just have a short discussion, because I mean, I think it's important just to reflect on this.

So the flu vaccine does change every year, or it has the potential to change every year, and so, you know, then the recommendations also have the potential to change every year.

And as was mentioned, I think, by the chair earlier, you know, there's now this new recommendation relating to the nasal flu vaccine which, you know, affects this one and some of the

other measures I think.

So, you know, I view the ACIP as equivalent to the U.S. Preventive Services Task Force which is what, you know, I think Congress said that and very, you know, it's not exactly the same. I kind of agree with Steve in some ways.

But it's very strong recommendations generally speaking, but, you know, they're tempered by kind of the real world and what's going on.

So, you know, I think we have to just reflect that and say that you know, in general the evidence is very strong, but then, you know, year to year, there might be some issues that affect measures, affect what people do, affect the trends, everything.

So, and you know, ongoing research is always ongoing also each flu season. So, and that's how they find those things out. So, you know, I'm fine with that, but it just seems like we should discuss it and, you know, hear how

maybe the measure developer deals with it, you know, because it seems like it's a little bit of a moving target.

But again I think it's important, but I don't think -- you know, the evidence doesn't seem like it's going to go in the negative direction, but I would never, you know, foresee that I guess myself.

So I think, you know, it's always, that's why I wanted to discuss it, really. And I do think it applies generally to all the measures that have flu in it, and so it's worth having a short discussion.

MS. MUNTHALI: Mary.

DR. BARTON: If I might, this is Mary Barton. I would just like to speak for NQF here, because NQF uses a process that asks us to update specifications or changes to specifications annually, where new guidance, there's a pathway for new guidance to be surfaced in a very timely manner.

And so I think while the committee

sees the endorsement that is coming at a several-year interval, in fact, the NQF staff are working with measure developers on a much more frequent basis.

And so I think, while I can understand your concerns I suspect if you were involved in some of the interim calls, you might feel more confident about how NQF is using the measure maintenance cycle to make sure that when codes change or practice shifts that those things are taken account of.

MS. MUNTHALI: And that's part of our annual update process. Thank you, Mary.

And if there are no material changes, you're likely not going to see them, but a material change would trigger an ad hoc review, and then we'd bring it in front of the committee.

So it is an annual review that's in between the three-year review, so there's an opportunity for developers to update their measures according --

MEMBER SALIVE: So I think to me, the

nasal flu recommendation is a material change, right, and I appreciate that there is this process.

I think I was part of one of the reviews in the past where things were being updated; maybe it was pneumococcal. But, so I appreciate that but I think, you know -- okay, enough said.

MS. MUNTHALI: And I think the point you raised about a moving target is something we should think about, because when we put out this call for measures, it was last year in October.

So developers are working with us throughout the process. I don't know if they will ever get caught up to whatever is most current, and you know, measure development is very resource intensive in terms of the dollars and staff time, budgeting.

So as much as possible we do know that what we're trying to do here, what we were asked to do by the federal government in 2008, is standardize immunization specifications. So we

ask that at a minimum, the measures that come through our process, whether they're influenza vaccine measures or pneumococcal vaccine measures that they're at least standardized in that sense.

But, you know, it's a point well taken. I don't know how we'll ever get caught up depending on, you know, where we are in our process and where developers are in their update process. They also have updating cycles as well.

CHAIR McINERNY: So for this year and future years, will you not count intranasal flu? Do you --

DR. BARTON: This is an adult measure.

CHAIR McINERNY: Right.

DR. BARTON: So there are no children in this measure.

CHAIR McINERNY: So, but what does ACIP say about intranasal flu for adults? What? Anybody know?

DR. BARTON: I believe it's applicable.

MEMBER TEUTSCH: Yes, but along the

same line, I'm really interested in how things are done when you get to variations, such as for the over-65 population, there's been an intent to use a higher potency vaccine. And as you know, there are age limits for the intranasal and they're likely to, they may well change it at various times.

To what extent does this measure actually capture that level of detail? And maybe even should it? I mean, if you got vaccinated with a regular vaccine and you're over 65, you know, you're getting partial protection, maybe even most of the protection, but you're not doing the optimal. So what do you capture, and how does it vary over time?

MS. WILLIAMS-BADER: So this measure as I pointed out, is collected using a CAHPS survey question, so we need to ask the question in a way that people are going to understand.

The question right now is have you had either a flu shot or flu spray in the nose since July 1st of the year in which we're measuring.

If we were to use claims, we would likely miss quite a few flu shots that were administered because the flu vaccines are being given in a number of different areas outside of the physician's office, and those claims are not necessarily getting to the plan.

So as a plan-level measure, we have to make a choice about which data source is likely going to give us the best information, and for this one, we think it's the survey question.

MEMBER TEUTSCH: But we were talking about evidence generally, and I'm just curious to what extent we, you know, how precise do we want to be?

I mean, in many ways, what you're saying is very practical and makes a lot of sense, but in some sense, it's not optimal. And I'm just curious as, not just for this one but for the whole set of measures, what's being done, and how consistent is it? And is there some harmonization across all of this?

MS. MUNTHALI: It's a great question.

The committee will have to determine the degree to which we're not prescriptive about, you know, the precision of any of what we require in terms of evidence or the testing.

However, because NQF has done significant work, and this committee has done significant work, in trying to make sure that at least we've accepted the ACIP guidelines, and at least for influenza vaccines, that they're harmonized; the measure specifications for the measures that are in front of you are harmonized to the extent possible.

Now there may be reasons why they can't be. There are different data sources, different things like that. But that will be up to your discretion to decide.

So I know it's not easy. We didn't give you an answer, but these are the sort of things you should be talking about.

And just to add, in your preliminary analyses, what we did as staff is to point out where these are standardized with our standard

specifications and where they're not and where the misalignment is.

Are there any other questions on evidence? Michael, any questions or comments?

MEMBER BAER: I have nothing further. Thanks.

MS. MUNTHALI: Okay, great. So we will vote on evidence for Measure 0039: Flu Vaccines for Adults ages 18 and older. One is high, two is moderate, three is low, and four is insufficient, and voting is open.

(Voting.)

MS. MUNTHALI: So 11 for high, 1 for moderate, and 3 -- I'm sorry. It switched over. I guess it was one. 11 high, 1 for moderate, and 3 low. So in percentages 85 percent high, 8 percent moderate, and 8 percent low.

And so we'll go to performance gap.

CHAIR McINERNY: By the way I just looked up the ACIP recommendation, and it says that the ACIP voted that live attenuated influenza vaccine, also known as nasal flu,

should not be used during the 2016-2017 seasons.

ACIP continues to recommend annual flu vaccine but needed to be inactivated influenza vaccine or recombinant influenza vaccine for everyone 6 months and older. So it sounds to me like that includes adults.

But I agree that -- then the next paragraph they go on and talk about the problem with LAIV for children through 17. I don't know, and it doesn't say anything about adults as having a problem, but the recommendation sounds like it should be for all.

So we probably ought to try and confirm that with ACIP, because I think it's important. And, you know, then the problem is with it using a CAHPS measure. Yes, you know, people are going to say, yes, I got a flu vaccine. Whether they say it's nasal or shot, you're not going to know.

DR. BARTON: Yes. Unfortunately the CAHPS cycle is such that this recent guidance would have been impossible to get into the

government approval for changing a CAHPS item.

It's somewhat ironic, but I think it's a good impetus to us as we continue to try and update our immunization measures in particular by looking at alternative data sources so that we're not reliant on survey.

CHAIR McINERNY: Right.

MS. MUNTHALI: Great. Catherine. Michael.

MEMBER HILL: All right, so now we're going to talk about the performance gap, and here we're looking at the fact that there is data to demonstrate some variation or overall less than optimal performance across providers and populations.

What we see submitted is that commercial mean is 49.2. The Medicaid mean is 39.8. In terms of disparities, here's another example, just like what Matt said earlier where we would like, the committee would like to see disparities assessed in order to target the populations, at-risk populations.

We do have some data that shows that influenza coverage was 31.5 percent among adults age 19 to 49, and 47.7 percent among adults age 50 to 67. That there were disparities in coverage observed for most racial and ethnic groups.

That our influenza coverage for whites age 19 and older was 47.6, versus that for blacks which was 36.5 percent, and for Hispanics, 33.2 percent.

The question before the committee was is there a gap in care that warrants a national performance measure, and I'd like to remind you of my point of view, and that is each year, we start with a hundred percent gap. And so it is an annual challenge that we have to drive, and it is really important.

This has been a critical measure for the 16 counties that I have practiced in and worked with to reduce hospitalizations and to regionally benchmark especially our rural health efforts where access to care is sometimes

challenging.

So I think it's really helpful to collect this data by this age category because we've been able to make some improvements in my area of the country by working with our existing infrastructure and processes. So it's rated high as an opportunity for improvement.

MS. MUNTHALI: Michael, anything to add?

MEMBER BAER: Not a whole lot to add. I mean, I agree with everything that Catherine said, and I'm just not sure how the CAHPS is administered and if it gathers, it doesn't sound like it gathers the ethnic, race, language, and I do feel this is important.

I thought there was an interesting comment that was on the preliminary evaluations about anti-vaxxers versus, well, anti-vaxxers maybe skewing or changing the disparities over time because of, you know, folks not wanting their children to get the flu shot.

And that predominately is in the

educated white population, whereas the less well-to-do families and those in the black and Hispanic populations would not have as high a rate.

So it's interesting how there could be an influence on this. I don't know how much of an influence that will be, but I do think that you know, it's one of those sub-groups that could affect disparities.

MS. MUNTHALI: Mary or Jenna, did you want to address --

DR. BARTON: No. I was just trying to get my head around appreciating lower rates in whites because it equals disparities, because that seems rather counterintuitive to me about, you know, what the health care system is, the responsibility of health care system to give everybody the vaccines that they need, everybody up to date.

So in terms of CAHPS, we're just trying to track down now the degree to which CAHPS data may be reported. I'm not sure that

race and ethnicity is fully included. Certainly it's not in the data that we get for the measure, which is again at a health plan level.

MEMBER BAER: Yes. And I think you've mentioned before that, you know, and I'm with a health plan, so at the health plan level I know that we do look at this.

You know, we have a Health Equities Council, and, you know, we certainly are going to be drilling down into our data. Not only do we drill down, but as we are a managed care company in Pennsylvania that has to answer to the administrator of the Medicaid program in Pennsylvania, the administrator of the program in Pennsylvania is keen on, you know, us looking at the disparities.

So, you know, while we may not get this at the CAHPS level, I do know that plan-level data is being drilled down into -- you know, in trying to be able to capture the race and ethnicity correctly for those who are in our plan is an issue. So I can see that that could

become an issue, you know, if we tried to do this at the CAHPS level.

So, you know, I do think that we struggle with trying to capture the correct race and ethnicity, but I do think that if it's not at the CAHPS level that we are looking at it at the plan level.

MS. MUNTHALI: Jacki, Matt, then Barry-Lewis.

MEMBER MOLINE: Just a brief question and comment. You know, there's so many folks that are getting their shots outside of the traditional medical setting that I wonder how much of it is actually being captured.

I know that when I see a patient I'm being asked if they've had the flu shot, and then if I'm not administering it, to record without ordering. But I don't know how many people are actually doing it, because it's a bit wonky to do that rather than just order it and have it administered that day.

But how many folks, how many people

are we missing in this, because you go to any CVS or any drugstore now, and it says flu shots are available, so these data are going to be missed because the health plans are not administering it. And unless someone comes in during the flu shot time for an evaluation, then the data will be missed because they've gotten it elsewhere.

MEMBER HILL: Indeed, we're seeing a lot of free vaccines done in North Texas, and a lot of employees who are getting free vaccines, and so it's just not accessible through administrative data the way you might think.

So patient's report is almost where you end up just by virtue of the fact that we have made it so accessible. We even have RNs in Texas who can receive lots of vaccine and vaccinate in their community centers, and we see that happening with no administrative billing records.

MS. WILLIAMS-BADER: And I would just say that's exactly why we use the survey question because it is, if they are getting it in any of

those non-physician settings, the measure is going to capture that.

MS. MUNTHALI: Matt.

MEMBER STIEFEL: Sorry to beat this dead horse of race and ethnicity, but it's a little different in this case because we're talking about CAHPS as opposed to HEDIS.

The rationale for HEDIS is that plans don't collect this information. For CAHPS, this is a situation where there are known disparities by race and ethnic groups, and it's -- the issue for CAHPS is that NCQA doesn't gather that information.

So I think the recommendation should even be stronger that for CAHPS, that that information is valuable.

MS. MUNTHALI: Barry-Lewis.

MEMBER HARRIS: I just wanted to make a comment that it's not difficult for me to wrap my mind around a disparity of what is considered a majority population, particularly with the area that I work in in Tennessee, Mississippi and

Arkansas.

So it just depends upon where you are and then the SES of that particular area, so it may not necessarily be what may be seen as mainstream that could have a disparity, because disparities exist wherever there is a gap in care, no matter who it is.

And so we should always think about just the, in population health, who is not receiving the care so we can reach whoever they are.

MS. MUNTHALI: Great. Other comments from the committee?

I think we're ready to vote on performance gap for Measure 0039. This, one is high, two is moderate, three is low and four is insufficient, and we're looking for 14 votes. Thanks. Thirteen, somebody stepped out. Actually 12.

(Voting.)

MS. MUNTHALI: So for high, 11 voted high; 1 voted moderate, so this measure passes on

performance gap. So we'll move on to reliability. Catherine.

MEMBER HILL: So now we're going to look at reliability, which requires that the measure produce consistent reliable and credible results about the quality of care when implemented, the level of analysis at the health plan level and integrated delivery system level.

The numerator here is those aged 18 to 64 of the Medicare or commercial CAHPS survey who report having received an influenza vaccine since July of the previous year, and the respondents 65-plus years to the Medicare CAHPS survey who report having received an influenza vaccine since July of the previous year.

The denominator is Medicaid commercial CAHPS respondents and Medicare CAHPS respondents. The developer notes a change in measure specifications for both groups where they had changed the wording from "have you had a flu shot since September 1st" of a particular year to "have you had either a flu shot or flu spray in

the nose since July 1st" of the specified year.

The developer also noted changes to the measure specifications for the younger age group and expanded the age range from 50 to 64 to 18 to 64 to align with the current ACIP guidelines and added Medicaid product line to the eligible population.

Both of these changes were reviewed and vetted through a public comment period and approved by the Committee on Performance Measurement and Board of Directors.

The question before the committee is are all the data elements clearly defined; are all appropriate codes included; and is this likely to be consistently implemented?

MS. MUNTHALI: Thanks, Catherine. Michael, anything to add?

MEMBER BAER: I have nothing to add. I mean if you wanted me to -- you know, as far as those questions are concerned, I can, you know, since it's a survey, we don't need to have any codes, so I think the data elements are clearly

defined, and I think it's consistently implemented because it's a CAHPS survey.

So I think that, you know, would really lend itself to consistency, so that's all I have to comment on.

MS. MUNTHALI: So thank you. Because this is a maintenance measure, the information, the new testing information NCQA supplied for reliability is directionally the same as it was when they brought the measure up for maintenance in 2012.

So the committee can opt not to re-vote and just accept your last decision on the measure with regards to reliability. Are you okay with that? Heads are shaking yes.

So we'll move on to validity, and Catherine and Michael, anything you'd like to mention for the group?

MEMBER BAER: I would like to just comment, Catherine, if you don't mind. You know, when they talked about validity testing and performing the empirical test on stakeholder

volunteers, I didn't know if the definition of volunteers could be clarified.

Were these folks who were representative of the populations being administered the survey, or who were they?

MS. MUNTHALI: Mary and Jenna.

DR. BARTON: Sure. So we endeavor to include in our cohort -- when we suggest a change to the CAHPS consortium, just so you could appreciate the steps here, we do the testing in a population absolutely that is representative of the population who are answering the survey, and we're working with endeavoring to match reading level, age group kinds of comprehension issues, and that is the process that we went through in this change.

I think that the bigger question, going back to both what Tom and Steve were talking about, is that there are guideline changes that come out at the time when FluMist was first recommended for adults, you know, then the folks sort of scurry to update the wording

and get the CAHPS consortium to include an updated question.

And now we're, you know, we're in a situation where CDC is saying, oh, well, maybe not this year. Don't do it this year.

So that's, you know, to the point of the process that we used, we used what we understand to be all of the best techniques for cognitive testing of the item.

MEMBER BAER: Great. You've answered my question, and I appreciate that. In fact, I'd rather see the, you know, keeping the language the way you have it now as, you know, if we happen to have a change in the recommendation next year and it does include the nasal spray, then you've already got it there, and you don't need to add it, so thank you.

MS. MUNTHALI: Other comments on validity?

MEMBER HILL: And so there are no meaningful trends in the missing data on this? It looked like there was one committee comment.

It was concerned about missing data.

MEMBER BAER: I can just -- I don't know if that was my response.

MEMBER HILL: Yes.

MEMBER BAER: I don't know if I can interject.

MEMBER HILL: Yes.

MEMBER BAER: I think we talked about ethnicity and race, you know, being not collected and potentially being missing data, but I think we've already discussed that. So I didn't have anything further on it.

MS. MUNTHALI: I think Mary and Jenna were looking up something, or is that --

MS. WILLIAMS-BADER: Well, I think we were just trying to figure out exactly what the question was about the missing data.

MS. MUNTHALI: That there wasn't any information on it, and so they wanted to know if there was any significant impact of there not being any information on the missing data.

MS. WILLIAMS-BADER: We don't get

information about missing data per rate.

MS. MUNTHALI: Any other questions?

CHAIR McINERNY: By the way, I did pull up a CAHPS survey,/k. and it does ask at the end of the survey about race and ethnicity. Now it's possible that the person who fills out the survey may not answer that question or those questions, but it does ask at least.

MEMBER STIEFEL: And does NCQA look at that?

DR. BARTON: We don't currently get it, but I think we can ask for it. And my understanding is that it is variably filled out, not highly complete, but there's no reason why we couldn't try to correlate the race and ethnicity that is reported with the response to the vaccination questions.

MS. MUNTHALI: Okay. It doesn't look like there are other comments or questions so we'll go ahead and vote on validity. One is high, two is moderate, three is low and four is insufficient.

(Voting.)

MS. MUNTHALI: It looks like we need four more votes, so if you can re-vote, point your clicker to Sheila or Diane.

CHAIR McINERNY: We're in Chicago; you have to vote often.

MS. MUNTHALI: So we're at 13. So 6 voted high; 7 voted moderate, so the measure passes on validity. We will go on to feasibility, so Catherine and Michael.

MEMBER HILL: On feasibility, the question is to what extent do the specifications including measure logic require data that are readily available or could be captured without undue burden and can be implemented for performance improvement?

The developer reports that the CAHPS survey is conducted by third-party vendors via telephone, mail, email or mixed protocols and that there is concern that many Medicare beneficiaries do not have access to a computer or internet to complete the survey in electronic

format.

There's also a concern that moving to an internet-based mode of administration would bias the results as older, more frail adults may be less likely to complete the survey. Some data elements are in defined electronic fields for electronic source.

The question before the committee is what is the burden of data collection for this measure, and the preliminary rating for feasibility is moderate.

MS. MUNTHALI: Michael, anything to add?

MEMBER BAER: No. I have nothing further to add. Thanks.

MS. MUNTHALI: Comments from the committee?

I think we're ready to vote on feasibility for 0039. One is high, two is moderate, three is low, and four is insufficient.

(Voting.)

MS. MUNTHALI: For feasibility, 6

voted high; 7 voted moderate, so this measure passes on feasibility, and we'll move on to usability and use. Catherine and Michael.

MEMBER HILL: All right. So usability and use is where we're looking at accountability and transparency. And let's see what I've got here.

There are related measures, as we have implied in earlier conversations, such as preventive care screening, influenza immunization in the end stage renal disease population and other flu measures, as well as the percent of residents assessed and appropriately given the seasonal vaccine.

The developer has noted that this measure is not completely harmonized with other related measures, as this measure is the only measure to collect information through a patient survey, and they don't view this measure as competing with other measures because of its source of a survey.

It has been observed that the other

measures are complementary to each other, and each of course confers some protection from getting influenza.

The 2012 NQF committee had suggested a universal measure that incorporated all of the various populations included in the influenza immunization measure. That's all I've got to offer.

MS. MUNTHALI: Great. Any comments?

MEMBER BAER: I'm Mike. No, I did not have any further comments.

MEMBER BIALEK: I have a question about the impact of the measure as a result of the plans. So in the past, most of the vaccinations would have been provided by the plans, and as mentioned earlier, CVS and others are providing it.

So are there any data that you provided that show patient X not vaccinated at the time originally coming in and later on vaccinated as a result of the plan?

So I'm wondering if the plan, in terms

of usability, are we measuring the impact of the plan, or are we measuring just the individual who chooses to become vaccinated anywhere? Like is there a way to know the impact of the plan on the individual?

MS. WILLIAMS-BADER: I would say no, there isn't a way to know what the impact is, but the plans certainly have a role to play in helping members to get vaccinated. They have different tools at their disposal to make sure that members are getting vaccinated.

So just like with any of our other plan-level measures or even physician-level measures and other measures of looking at different levels of accountability, I'd say it's hard to tease out exactly what is, you know, what exactly is the cause and effect of the plan getting the member vaccinated.

But we do at least think that this is within a plan's control to impact the rate, and that's what is most important is that they can actually impact the rate.

MEMBER HILL: So from a harmonization perspective, what is our risk of over-measuring or measuring someone multiple times without really having an advantage, since there are additional, you know, other sources of this information being collected, related or competing measures?

MS. WILLIAMS-BADER: Well, again, because this is a health plan-level measure, it's a population-level measure, I would say one of the advantages that it has or a way it complements other measures is that it's going to -- that plans can reach out to members who are not coming in and seeing physicians.

So if you have a physician-level measure, physicians certainly can have an impact but only if the patient comes in. Whereas a health plan-level measure, they can be reaching out to their members who aren't interacting in other ways with the health care system and reminding them about the importance of getting flu vaccination and perhaps pointing out to them

where they can get vaccinated.

DR. BARTON: And I would just want to add that in looking at the other measures, the population, the general population who are indicated for a flu shot is so much larger than most of those measures that apply to specific settings that it's hard to imagine how, you know, measuring only the ESRD population would be an adequate way to assess a health plan's responsibility to their members.

MEMBER HILL: I was noticing on the 0227 it just says influenza immunization.

MS. MUNTHALI: So I'm glad you mentioned harmonization and related and competing measures. So what we're asking the committee to do is evaluate each measure that's in front of you on its own merits.

And on day two, tomorrow, at the end of the day after we've evaluated all of the measures that might be related or competing, we will then look at the differences and, you know, talk about data source and all of that.

So we're just looking at this measure for now, evaluating it on its scientific merits.

MEMBER HILL: Thank you.

MEMBER BAER: So I'd just like to make a comment on something that has just been mentioned about the plan affecting the flu shot rate and getting the patient into care.

I think the biggest, the one, the biggest single influencer of someone getting the flu shot is the physician advising the patient to get the flu shot.

So the plan can get the patient to the appointment, but unless the doctor's advising the patient to get the flu shot, it may not be in the plan's control to get that patient the flu shot. So I'll just put that out there that, you know, that there is a physician responsibility for advising the patient to get the flu shot.

MS. MUNTHALI: Thanks, Michael. Any other comments before we vote?

Okay, so for measure 0039, usability and use, high, one; two, moderate; three, low;

four, insufficient information.

(Voting.)

MS. MUNTHALI: Okay, so 6 voted high; 7 voted moderate, so this measure passes on usability and use. For overall suitability for endorsement your options are one, yes; two, no, and Sheila's queuing up the slides.

(Voting.)

MS. MUNTHALI: More votes, one more if you can click again. So 12 yes, 1 no. This measure is recommended for endorsement. Thank you all.

What we're going to do is change the schedule a little bit. Lunch is here but we're behind, so, but you did very well, but we're still behind. We have a lot of measures in front of us.

So what we're going to recommend is that we have a working lunch for about 15 minutes. We ask that you get your food and you can sit in the back if there's room, but definitely sit here or in the hallway. And we

reconvene at 12:45.

But before we do that, we have public and member comments, so we want to see if there are any of our members or members of the public on the phone that would like to say something, or anyone in the back of the room.

OPERATOR: Thank you. At this time, if you'd like to make a comment, please press star then the number 1 on your telephone keypad. And there are no public comments over the phone at this time.

MS. MUNTHALI: There are no comments in the back, so we can break for lunch. 12:45 we'll be back. Thanks.

(Whereupon, the above-entitled matter went off the record at 12:25 p.m. and resumed at 12:48 p.m.)

MS. MUNTHALI: Hi everyone. We're going to get started.

CHAIR McINERNY: Thank you, everyone, for reporting back promptly, and we'll continue now on our working lunch with consideration of

our next maintenance candidate measure, 000226.

MS. MUNTHALI: And so, Operator, I just wanted to make sure that Lisa McGonigal's line is open.

OPERATOR: It is.

MS. MUNTHALI: Hi, Lisa.

DR. MCGONIGAL: Yes, hi. Can you guys hear me okay?

MS. MUNTHALI: We can hear you just fine.

DR. MCGONIGAL: Great, thank you.

MS. MUNTHALI: Lisa, can you just give a background on behalf of your developer team?

DR. MCGONIGAL: Yes, I will. Again I'm Lisa McGonigal, and thank you all for taking the time today to review this measure.

Again it is NQF's measure number 0226, Influenza Immunization in the End Stage Renal Disease Population which was developed by the Kidney Care Quality Alliance, or KCQA, so I'll just provide a very brief overview for you.

So unlike the other measures that

you've reviewed already today, this measure is specified for assessment at the level of the dialysis facility. It applies to all end stage renal disease patients aged 6 months and older, and as with the other flu measures that you've been discussing the measure is entirely consistent with NQF's standardized specifications for influenza vaccinations as well as with the current recommendations from the CDC Advisory Committee on Immunization Practices.

To illustrate the importance of the measure we note that infectious disease is the second leading cause of death among patients with ESRD, and pulmonary infectious mortality including influenza related deaths is tenfold higher in the ESRD population than in the general population.

Yet despite this and the longstanding guidelines and recommendations in place that this vulnerable population be routinely immunized, data from our major testing and the most recent United States Renal Data System Report indicate

that there is a persistent and substantial performance gap with only 71 percent of ESRD patients receiving the vaccine in the 2012 to '13 flu season and there's wide variation in facility performance scores on the measure ranging from 78 to 100 percent, both indicating that there's still substantial room for improvement in this aspect of care.

In regards to the scientific acceptability of the measure, testing was not redone because the measure is a maintenance measure. However, we'd note that during its last endorsement maintenance review in 2012, the committee rated the measure reliability as high and the validity as moderate.

In regards to feasibility and usability, the measure is currently being used for internal quality improvement in dialysis organizations.

Additionally, in its proposed rule for the ESRD Quality Incentive Program issued in June of this year, CMS indicated that it's seeking to

add an influenza vaccination measure to the program in the future and currently ours is the only NQF endorsed ESRD flu immunization measure that would satisfy this requirement for use in the QIP.

The ACIP has also been in discussions with CMS regarding an update to build the necessary data elements into their CROWNWeb data repository system, so it appears and we believe that measure will be incorporated into the program in an upcoming cycle.

So finally, we also reviewed the standing committee pre-meeting review evaluations and I just wanted to comment on one that I haven't already addressed in the introduction. This is on the addition of pediatric patients in measure testing or the lack thereof.

So the measure received time limited endorsement in 2007 as an adult-only measure, then in 2008 NQF released its standardized specifications for influenza vaccination which included children.

1 And in response to this and on the
2 recommendation to the American Society of
3 Pediatric Nephrology, we did change our
4 specifications but we did not retest the entire
5 measure.
6 We note that date of birth is a
7 standard data field that would not materially
8 affect data collection or testing results, and we
9 did not believe that confirmation of reliability
10 and validity with the expanded applicable age
11 range was warranted.
12 And that's it. Just let me know if
13 you have any questions as you discuss.
14 MS. MUNTHALI: Thank you, Lisa. So I
15 will pose the same suggestion I posed earlier
16 when we were discussing the NCQA influenza
17 measure on whether or not the committee wants to
18 just vote en bloc for the remaining influenza
19 measures on evidence only and then continue with
20 your discussion on performance gap, or in this
21 case since it is a maintenance measure you can
22 opt to just take your recommendation from 2012

when the measure was last looked at.

So any, no discussion on the evidence given the lengthy discussion we had last time? Okay, everyone's -- well, a lot of people are shaking their heads yes, so we're going to proceed to performance gap. And I think it's Ron and Tom.

MEMBER BIALEK: Well, as the measure developer mentioned there is a performance gap of, I think it was 78 percent, 71 percent of the population immunized and demonstrated a gap amongst facilities as well. So the evidence seemed pretty straightforward on that.

CHAIR McINERNY: I don't have anything to add, pretty straightforward.

MS. MUNTHALI: Great. Other comments? Okay, it looks like we can vote on performance gap for measure 0226. We're pulling up the voting slides, so this is for performance gap.

Okay, so performance gap, 1 high, 2 moderate, 3 low, 4 insufficient.

(Voting.)

MS. MUNTHALI: And we need one more vote, so if you can re-click. We have 13. So five voted high, seven voted moderate and one low so this measure passes on performance gap, so we can go into our discussion on reliability.

Ron.

MEMBER BIALEK: So the data were specified well for both numerator and denominator. The one question I had and the measure developer mentioned it is that for the testing there were no pediatric data.

And I, while the measure developer said they would not think there would be a difference in the recording of and collection of the pediatric data, I didn't know if that's really valid.

That was really my main question about the, because the measure includes 6 months and over.

MS. MUNTHALI: Lisa.

DR. MCGONIGAL: Yes, did you want me to comment?

MS. MUNTHALI: Yes.

DR. MCGONIGAL: Yes. The measure was tested at the level of the data elements and as I noted the date of birth field is a common standard data element so we did not believe that there would be any issues at all in capturing this reliably.

And that it is already known as a valid data element so it would not impact our testing results. As to how it would impact the performance on the measure that's a different story, but it would not impact reliability or validity of the measure.

MS. MUNTHALI: Ron.

MEMBER BIALEK: Well, the data for the pediatric population would be provided by a guardian or a parent, correct, whereas the data for the other, the adult population we would confirm with the patient themselves, right.

So it would seem like there could be some difference, some discrepancy there.

MS. MUNTHALI: Lisa, did you --

DR. MCGONIGAL: I'm sorry.

MS. MUNTHALI: Go ahead.

DR. MCGONIGAL: It would depend on whether, if the vaccine is administered at the facility then that is not an issue. But otherwise yes, it would be in many cases confirmed by a patient or guardian.

MS. MUNTHALI: Tom, would you like to add anything?

CHAIR McINERNY: No.

MS. MUNTHALI: Okay. Any other comments on reliability testing for 0226?

DR. MCGONIGAL: Oh, can I add one additional thing? I'm sorry.

MS. MUNTHALI: Sure.

DR. MCGONIGAL: Okay. Yes, I just also wanted to add that as far as our experience with other measures in the population that pediatric, it's fairly negligible in the standard. The population is negligible in the standard facility, meaning non-peds based facility.

And in fact most facilities, actually I don't think there have been any facilities that have met the CMS threshold of greater than 11 pediatric patients to be included in the measure if that's, and that's at this point in time.

According to USRDS there are about fewer than 10,000 children being treated with ESRDS.

MS. MUNTHALI: Ron.

MEMBER BIALEK: Was the gap any greater in the facilities who have a higher proportion of the pediatric population?

DR. MCGONIGAL: Are you referring to the gap from testing or from USRDS data?

MEMBER BIALEK: I'm sorry, the gap in testing.

DR. MCGONIGAL: Yes, we did not test in the pediatric patients so we don't have that information.

MEMBER BIALEK: Right, but I'm talking about the -- I'm sorry. The facilities where you show a higher proportion of non-vaccinated or not

ask the question was that any greater for facilities that had a higher proportion of pediatric population?

DR. MCGONIGAL: No, we did not assess that during testing.

MS. MUNTHALI: Other questions or comments?

Okay, we're ready for a vote for 0226, reliability. 1 high, 2 moderate, 3 low and 4 insufficient, voting is open. We're looking for 14 votes.

(Voting.)

MS. MUNTHALI: Okay, one high, 11 voted moderate, one voted low and one insufficient, so this measure passes for reliability and we'll move on to our discussion in validity. Ron and Tom.

MEMBER BIALEK: So I had the, really the same issue about the pediatric population when it came to validity testing as well, because that population again was excluded from testing.

MS. MUNTHALI: Tom, anything to add?

CHAIR McINERNY: No, I agree. It would be helpful in the coming years to try and look at what the pediatric population results are as well as the adult and I think there should be ways of doing that. And I would urge the measure developer to look into that please.

MS. MUNTHALI: Lisa.

DR. MCGONIGAL: Thank you. Yes, and we do want to reinforce again that the testing was performed at the level of the data elements and it's the date of birth data element that we're looking at for pediatrics which is a common data field and should present no issues as far as reliability and validity are concerned.

MS. MUNTHALI: Other comments?

Okay, I think we're ready for a vote on validity. 1 is high, 2 moderate, 3 low, 4 insufficient. Voting is open.

(Voting.)

MS. MUNTHALI: We need one more vote. You can reselect. One more time. If you can try it one more time. Okay, 14. Zero voted high, 13

voted moderate and one voted no, so this measure passes validity. We'll go to feasibility.

So Ron and Tom.

MEMBER BIALEK: Again what was provided by the measure developer indicated that these data are routinely collected and seem to be feasible to collect.

MS. MUNTHALI: Any other comments?

I think we're ready for a vote. 1 high, 2 moderate, 3 low, 4 insufficient.

(Voting.)

MS. MUNTHALI: We need one more vote, if you can re-enter your selection. We got it, okay. Six voted high, eight voted moderate, so this measure passes feasibility and so we'll move on to usability and use.

MEMBER BIALEK: Yes. Currently being used, identifying gaps, quality improvement can be sold from this as well.

MS. MUNTHALI: Okay, great. Any concerns from the rest of the committee, comments?

I think we're ready to vote. 1 high, 2 moderate, 3 low and 4 insufficient information.

(Voting.)

MS. MUNTHALI: We got it, 14. Nine voted high, five voted moderate, so this measure passes on usability and use. So now we'll assess the overall suitability for endorsement. 1 is yes and 2 is no.

(Voting.)

MS. MUNTHALI: We need three more votes.

(Voting.)

MS. MUNTHALI: Two more. One more.

(Voting.)

MS. MUNTHALI: Just one more. We got it. So 13 voted yes and one voted no, so measure 0226 is recommended for endorsement.

Thank you, Lisa.

DR. MCGONIGAL: Okay, thank you so much.

MS. MUNTHALI: So now we'll call up our colleagues from the CDC for measure 0431:

Influenza Vaccination Coverage Among Healthcare Personnel. So I don't know if they're, are they in person or on the phone?

MS. LINDLEY: Hi, this is Megan Lindley from CDC. I'm on the phone.

MS. MUNTHALI: Hi, Megan. So if you could please give us a two- to three-minute intro of your measure that would be great.

MS. LINDLEY: Oh, sure. So this is a facility level measure looking at seasonal influenza vaccination among health care personnel and that's divided into three groups.

The first are payroll employees, the second are non-employee licensed independent practitioners whom we define as physicians, nurses in advanced practice and physician assistants, and the third group is non-employees also. It's students and health care trainees and volunteers age 18 and older.

All personnel who work in a reporting facility physically during the defined influenza season which is October 1st through March 31st of

the following year are included in the measure, so there's no exclusion based on clinical responsibility or patient contact.

The numerator categories are vaccination at the facility or outside, medical contraindication, declination and unknown status. This is consistent with the NQF harmonized -- excuse me -- consensus standards on vaccination with the exception of that unknown status category which we added to assist facilities in tracking their ability to report.

So this measure was last reviewed and endorsed by NQF in May 2012. Since the last endorsements we did make one change. We expanded the denominator. It used to be personnel working 30 days or more during the influenza season. It's now personnel working one day or more, so it's actually become more inclusive.

And this is based primarily on feedback from facilities regarding feasibility of identifying personnel working 30 days or more, and this is consistent with the specification

that we pilot tested so we were comfortable with the change.

Our data do show a gap in performance. Looking at the mean vaccination for all facility types there's room for improvement toward the Healthy People 2020 target of 90 percent vaccination.

The data also show substantial geographic variation. For just one example, the reported coverage in acute care hospitals in this past reporting year range from 63 percent to 97 percent, and then on the state end the variation is consistent across all the facility types that we've looked at and the data do also show variation among those three different reported groups of health care personnel.

We did also see some progress for the facility types that have reported for multiple years. They show incremental increases in reported coverage and decreases in the proportion of personnel with unknown vaccination status.

And the measure is currently in use in

eight CMS quality reporting programs which cover about 16,000 facilities. It began in January 2013 for acute care hospital inpatient quality reporting, inpatient rehabilitation facilities, long-term acute care hospital outpatient departments and ambulatory surgery were added in the 2014-15 influenza season.

Outpatient dialysis facilities and inpatient psychiatric facilities were added this past season 2015-16, and the PPS-exempt cancer hospitals will be added beginning in the 2016-17 influenza season, so the one we're in right now. Thank you.

MS. MUNTHALI: Thanks, Megan. As with the other influenza measures just wanted to pose a couple of questions to the committee. Do you want to accept the evidence that you reviewed, or the Health and Well-Being Committee reviewed in 2012, or would you like to have discussion and vote on this measure in particular?

Jason.

MEMBER SPANGLER: I just have a

question because -- and I was in here earlier. Sorry, I had to leave for another event. But I think this is the first measure that we're looking at where the dates are present. Is that correct, today, the October to March time frame we're talking about influenza season? Is that right?

MS. MUNTHALI: I think the other ones do because those would comport with our standard specifications. Not all of them though, but most of them, yes. Yes.

MEMBER SPANGLER: Oh. So I'm just wondering, because like a measure like this, if somebody got their vaccination on April 2nd they would fall out, but do we really want that?

I mean do we, I know the evidence around the season and when we have influenza and stuff like that and trying to get a vaccination earlier, I mean, I feel like, you know, I remember a few years ago talking with people from the CDC and they were encouraging people to get vaccinated in August and September.

I mean, the earlier they can get the vaccine in -- what's that?

No, no, no. But the push was for earlier, earlier, earlier if they could get the vaccine developed. So I'm just wondering if, I'm just bringing up the conversation about the season.

If we want to keep the dates just because, you know, it's the same thing with this one. If someone got vaccinated in September, mid-September then they're not considered, and what the committee feels about that.

MS. LINDLEY: This is Megan. That's a great clarification and it's a frequent question by our facilities too, so I'm glad you brought it up.

The October 1st through March 31st time frame is for the denominator only. The vaccination is beginning as soon as vaccine becomes available for the season, so somebody who is vaccinated in August or September depending on availability would be included.

The denominator is fixed in that way to sort define the population and account for potential delays in vaccine availability, but the numerator does allow for early vaccination.

You're correct that somebody vaccinated on April 2nd, the very end of the season would not be included because the cut-off for March 31st is the same for the numerator and denominator.

MEMBER SPANGLER: I'm sorry, can you clarify that because I'm looking at the numerator and it's referring to the denominator.

But where --

MS. LINDLEY: This is all personnel in the denominator who were vaccinated and vaccine became available through March 31st.

MEMBER SPANGLER: Oh, I see in the parentheses there.

MS. LINDLEY: So somebody who is vaccinated -- yes.

MEMBER SPANGLER: Or when -- okay, sorry. Thank you.

MS. LINDLEY: -- in September and then quit September 29th, they're not counted. Otherwise they are.

MS. MUNTHALI: Other comments or questions about evidence or anything that you'd like Megan to clarify?

Arjun.

MEMBER VENKATESH: I guess it's sort of the same question as Jason's as my read of this though is I see the or when the vaccine becomes available. My guess is that you construct the seasonality but recognizing that providers are going to move between facilities and move around as well.

So if you worked at a hospital that had access to vaccine and they gave a bunch of immunizations in September and then you go and you're a traveler and you work at a different hospital three months later, you're going to have a day of work in December and so you'll be captured in that facility or that hospital's IQR/OQR measure for this, but your immunization

happened at another hospital prior to the window that's, because my read of this is, you know, when was it available.

What if it was only available on October 1st at that facility? Would that mean that that score for that facility is wrong? Like do you need to have the dates on the numerator or could you just take it out?

MS. LINDLEY: No, if I'm understanding your question, so in the case of the traveler you cited we would encourage the facility, if there's sister facilities in the similar system we'll say they can use their data systems to pull the data and say the person is vaccinated at the facility.

Otherwise they would fall into the other vaccination category which is vaccinated outside the facility and provided documentation, and that documentation would be an attestation by the worker or a form from one facility or the other saying they were vaccinated.

So you don't require the date of vaccine availability at each individual facility

in order to score someone as having received the vaccine because they can be vaccinated within or outside the facility.

Did I understand your question correctly?

MEMBER SPANGLER: Yes, thanks. That helps.

MS. MUNTHALI: Steve.

MEMBER TEUTSCH: I also have a clarification question. It says in the numerator at least that you're included in the numerator if you decline influenza vaccination?

MS. LINDLEY: Yes, and I believe --

MEMBER TEUTSCH: I mean, it seems to me that you'd want these facilities to get those people. You wouldn't want to exclude them from the -- you don't want to include them in the numerators if they were vaccinated, or do I misunderstand?

MS. LINDLEY: No, no, no. And thank you for bringing up that clarification as well. This I also believe is consistent with the way

that NQF suggests that vaccination measurement numerators be constructed, so each category is available there for analysis.

But the way the data are actually recorded and scored it's only those personnel who receive the vaccine who would be counted in a compliance score and that's what CMS uses.

So they're measured, so you could calculate, for example, a declination rate, the declinations over the full denominator, but those people who are contraindicated, declined or unknown would not be counted in a performance score. They're not considered vaccinated obviously.

MS. MUNTHALI: And Megan is right, and I just wanted to add to that. In the numerator for our standard specs so we would include the number of persons in the denominator who received influenza vaccine or were assessed and offered but declined the vaccine or were assessed and determined to have had a medical contraindication of that.

And then how she says it's reported is different, but we want to make sure that you're assessing that throughout. And then -- does that make sense?

MEMBER TEUTSCH: I think it makes sense. It's just, I think it probably could be stated more clearly that you're collecting this information but you're going to actually be reporting it in ways that reflect these different categories, because it sort of sounded like they were going to be aggregated which obviously is not what you wanted.

MS. MUNTHALI: Any other questions?

MS. LINDLEY: This is Megan. Could I add just one thing for the committee's information?

MS. MUNTHALI: Sure.

MS. LINDLEY: I think there was a question earlier about the ACIP recommendations for intranasal vaccine but I don't know if it was resolved.

And I want to stress here that I'm

speaking individually and not on behalf of CDC, but I did find a sentence in, it'd be 2016-17 recommendations that says ACIP recommends that LAIV IV not be used during the 2016-17 season for any population. So those too extends to adults as well as children.

MS. MUNTHALI: Okay. I think there are a number of clarifying questions so I'm just going to recommend that the committee vote on evidence.

So Sheila, if you can pull up evidence. Okay. So 1 high, 2 moderate, 3 low, 4 insufficient, and we're looking for 14 votes, right.

(Voting.)

MS. MUNTHALI: One more. We got it. So five voted high, nine voted moderate so this measure passes on evidence, so we'll proceed to performance gap. Matt.

MEMBER STIEFEL: We're in our third of nine flu shot measures, so hopefully we're getting in a groove here.

So this is about opportunity for improvement in performance gaps. The developer noted already there are continuing significant performance gaps across types of facilities, across types of personnel and across geographies.

The data showed an upward trend for acute care hospitals, but still with remaining opportunity for improvement. The performance across different, the mean performance across the different types of facilities range from 76 to 88 percent and the standard deviation ranged from 15 to 23 percent.

Similar to the other discussions about measures, the disparities data aren't available because the data reported at the facility --

MS. MUNTHALI: Thank you. Patricia, anything to add?

MEMBER McKANE: No.

CHAIR McINERNY: So we're ready to vote on the gap for this measure, 0431.

MEMBER TEUTSCH: Yes. Could I ask one more question before we move to that?

CHAIR McINERNY: Yes.

MEMBER TEUTSCH: Matt, you know, it'd be interesting to stratify it by those three groups of different type of personnel. In particular, you care about the people who have patient contact because that's at least how I think of we are primarily trying to protect.

Do you have that data that shows the rates by those different groups, because if this were up at 98 percent for those who have contact it seems to me there would be relatively little opportunity for improvement.

MS. LINDLEY: We don't measure it by patient contact or clinical duties and the reason is that's not consistent with the ACIP recommendations.

The requirement for the denominator that all the personnel measured be physically present in the facility in performing a work duty is what we believe indicates their risk, because they do have the opportunity both to come into contact with patients or be in the patient's

room, for example, if you're talking about nutritional services, environmental services that kind of thing.

They also have the opportunity to come in contact with each other and transmit influenza that way. So we don't collect and we would not be able to stratify based on patient contact.

MEMBER TEUTSCH: No. And so I would think that that would be a useful thing to do, because there are some people who just have purely administrative functions and other kinds of things, and just think about it in terms of where the gaps really are.

CHAIR McINERNY: You could argue that the administrators should set the tone.

MS. MUNTHALI: It looks like we're ready for a vote on performance gap. 1 high, 2 moderate, 3 low, 4 insufficient.

(Voting.)

MS. MUNTHALI: So for performance gap for measure 0431 three voted high and 11 voted moderate so we'll move on to reliability. Matt

and Patricia.

MEMBER STIEFEL: So in terms of reliability the first question is about the clear specification of the data elements. I guess the only comment I would make is the one that Steve raised earlier about, because it was a little unclear when in the numerator definition it included people who had declined and that was confusing.

So I don't know if that's the developer's fault or not, but it would help to clarify that. Otherwise I thought that the numerator and denominator were clearly specified.

In terms of reliability testing, that's still part of this one. Yes. There were two types of reliability testing. One was interrater reliability where project staff compared to the raters from the facilities and in three jurisdictions, and the interrater reliability was quite high in two of the three and the third one wasn't as high but that may have been because the project staff didn't have

access to the full data from the facility.

And there were also key studies done where facilities received vignettes, case studies, and were asked to describe how they would react to different situations and how they would classify people. In both cases the reliability was shown to be high.

I guess one thing about the case studies, there were some problematic denominator elements including poor understanding of how to classify physician owners of health care facilities who worked part time and physicians who were credentialed by a facility but had not admitted patients in the past 12 months, and in the numerator some confusion about how to report persistent deferrals of vaccination and verbal declinations.

MS. MUNTHALI: Thanks, Matt. Patricia.

MEMBER McKANE: Yes. I think also one of the questions that we were asked, and I'm thinking this is all in the same section, was

that is assessed sample adequate for generalized or for widespread implementation?

And it was one of the areas that I thought that they did four different, and if I was reading this and understanding this correctly, there were four different sites that were chosen but there really wasn't much geographic variation.

And I'm not sure. I think we did a pretty good job about different types of providers, if I'm remembering correctly, and I've gotten these measures a little bit messed in my mind. But, and it's not something I would necessarily hold up or, you know, on this measure about, but I was just curious.

I thought typically we liked to try to get more geographic, because there was nobody from this, I think there was no Midwesterner. There was nobody from the Midwest, no facility from the Midwest or from the South.

And if that -- and so we're basically to generalize provider population we're assuming

that the geographic regions that were selected are representative of the general, of the nation as a whole.

MS. MUNTHALI: Jacki.

MEMBER MOLINE: There are also state mandates in certain states like New York, which was one of the four they chose, which require health care workers to be vaccinated or wear a mask.

California, I believe, has recently passed one or there are some mandates. They may be, L.A. County might have one but I'm not sure if it's throughout the whole state.

So I'm looking and seeing that the four states they chose, one of them definitely has had a mandate for at least three to four years or maybe more. One of them has a partial state coverage.

So it's also, I don't know if it's truly representative because the data will be skewed somewhat because of a state mandate for health care workers.

MS. MUNTHALI: Megan, would you like to address the geographic variation issues that have been raised?

MS. LINDLEY: Sure. I think it's correct that they're not necessarily geographically representative. Clearly this was a project where, and I believe this was discussed in our original submission to NQF, the selection of locations for participation was based on interest by the state and an ability to participate.

So it was not a scientific sampling, something that was done with no additional budget. Regarding which states were selected, California has a, it's not a strict mandate but they've had a health care personnel offering a documentation requirement, I believe, since 2006.

It's correct that New York now has a requirement, but at the time of our pilot testing which was on 2010 that requirement was not in force, so California is the state where you might expect the results to have been skewed or

additionally supplemented by the fact that they had a relatively recent requirement to track health care personnel vaccination.

We obviously have much more geographically representative data now that the measure is in use across the country. I think the challenge is that the reliability testing which requires in-person validation are looking at a bunch of records. It's extremely resource intensive.

MS. MUNTHALI: Thank you. Any other comments?

Okay, so we'll move on to a vote. So we should be on validity, right? Did we do reliability? Oh, reliability, sorry. Moving ahead. 1 high, 2 moderate, 3 low, 4 insufficient.

(Voting.)

MS. MUNTHALI: One person voted high, 13 voted moderate so this measure passes on reliability. So now we'll move on to validity and I'll ask Matt and Patricia to lead us in

discussion.

MEMBER STIEFEL: They also did two types of validity testing, convergent validity and face validity. For convergent validity, did kind of an interesting analysis of the correlation between the number of strategies employed to improve the rates, any improvement in the rate, and they found borderline significant, two significant associations between the number of strategies employed and improvement in rates.

For face validity they used a Delphi panel in 2011, really just assessing the appropriateness and clarity of the specification of the measure. And in that expert review in two rounds there was strong consensus on the specification of the measure.

Patricia, I don't know if you had anything to add.

MS. MUNTHALI: Thank you. Any other comments?

MEMBER STIEFEL: Oh, just maybe one on the convergent validity. I was intrigued by the

method and but though wondered, I suppose you could employ three or four bad strategies versus one good strategy, you know.

MS. LINDLEY: That's an excellent point. Just let me clarify that all the strategies that were surveyed and used in the analysis are the evidence-based strategies known to be associated with increased influenza vaccination. So we hope they're all good strategies.

MS. MUNTHALI: Okay. So I think we're ready for a vote on validity. 1 high, 2 moderate, 3 low, 4 insufficient.

(Voting.)

MS. MUNTHALI: Need two more votes. One more. Still need one more, if you can reselect. Sorry about that.

Okay, so three voted high, 11 voted moderate for validity for measure 0431. So we'll move on to feasibility. Matt and Patricia.

MEMBER STIEFEL: I'd just point out that the developer notes that because in many

cases clinicians and employees of facilities aren't necessarily part of the electronic medical record of that facility, it's more difficult to capture this information electronically so multiple modes need to be used.

And I think they noted some difficulty in documenting verbal declines of staff who verbally declined the immunization.

MEMBER McKANE: And I was also wondering when I was reading through this, in paper records and, you know, what the burden is on the facilities although this is currently, if I'm reading this correctly this is a current measure so this is actually being done.

So they are able to do it, but I was just wondering about the burden on facilities if that was, if there's any information about that from the developer.

MS. MUNTHALI: Megan.

MS. LINDLEY: Oh yes. I think what we have is anecdotal information on the burden, because as part of supporting the measure we have

a help desk so we're responding to TAs or queries.

We've certainly heard from some facilities and seen in our pilot testing and our published evaluation of the first year of hospital reporting that for large facilities with a lot of staff in some cases this can be burdensome.

We haven't received any more of what I would call large-scale burden information that they've really got this deluge of this measure is not possible when we have the 30-day requirement versus the one-day requirement in place for the denominator.

So I do think it is, it may be a burden for some facilities. It really is dependent on what kind of system the facility has used, because some of the larger facilities tend to have electronic records for their staff.

So it's difficult to say in a cohesive way what we know about burden, and I think there is also an extent to which the more the measure

is reported the easier it becomes.

So in case of hospitals which have the most employees out of all the reporting facilities, they've now been doing this since January 2013 so I think the burden probably is lessened.

MS. MUNTHALI: Thanks, Megan. Any other questions?

Okay, I think we can vote on feasibility, 1 high, 2 moderate, 3 low, 4 insufficient and we're looking for 14 votes.

(Voting.)

MS. MUNTHALI: We have 14. Two people voted high, 12 voted moderate and so this measure passes on feasibility. So now we'll assess the usability and use and turn it over to Matt and Patricia.

MEMBER STIEFEL: So in terms of the use of the measure it's in widespread use in a number of CMS and Joint Commission programs for facility accreditation and reporting. And we talked about the usability and use of the measure

in terms of the variation and the differences across quartiles in performance.

Let's see, what else did I have? And in terms of the use of the measure, I think they've been able to show especially in long-term care facilities a demonstrated association between the measure and improved patient morbidity and mortality, which is important and I think it is somewhat unique, more than can be said for a lot of measures where you can actually see a significant outcome improvement from the use of the measure, associated with the use of the measure.

MS. MUNTHALI: Thank you.

Patricia.

CHAIR McINERNY: I have a question. Does CMS put any teeth into this by having a disincentive for low rates or do they dock the hospital or other system as they do for some other measures now?

MS. LINDLEY: So at this time the measure is part of the quality reporting programs

in the pay-for-reporting aspect, so there's a significant disincentive to fail to report. It's a potential two percent decrease in the annual payment update from CMS.

None of the measures -- excuse me. None of the programs at this time include this in the value-based purchasing pay-for-performance, so at this time CMS hasn't specified a level of vaccination that needs to be obtained, only that it must be reported.

MS. MUNTHALI: Other questions?

Okay, we're ready for a vote on usability and use for measure 0431. 1 is high, 2 is moderate, 3 is low and 4 is insufficient information.

(Voting.)

MS. MUNTHALI: So 11 voted high and three voted moderate for usability and use for measure 0431. So we'll proceed to an overall suitability for endorsement vote. 1 is yes and 2 is no.

(Voting.)

MS. MUNTHALI: We need one more vote. Try it one more time. Okay. We have it. So it's unanimous, 14 voted yes for overall suitability for endorsement for measure 0431.

So Megan, thank you. Just one thing before we close with this measure. With regards to the NQF standardized specs, what we include and the numerator are that they can be computed and reported separately.

So perhaps in revising your measure you might want to add that to clarify some of the concerns that were raised by the committee.

MS. LINDLEY: Yes, thank you. And I think we can also specify what the performance score is based on in addition to the numerator element.

MS. MUNTHALI: Okay, great. Thank you.

MS. LINDLEY: Thank for the opportunity to join.

MS. MUNTHALI: Okay, so our next measure for review, it's another influenza

immunization measure. This one is from PCP Foundation. It's Measure 0041: Preventive Care and Screening Influenza Immunization. And our developers are in the room, so I'll ask them to come up.

MEMBER TEUTSCH: Okay, this is influenza continued.

MS. MUNTHALI: Steve, Steve, we're going to ask the developer to introduce the measure.

MEMBER TEUTSCH: Oh, I'm sorry.

MS. MUNTHALI: That's okay. He's excited.

MS. CHAVARRIA: I think he's saying Ron made him do it. So hello, everyone. My name is Elvia Chavarria. I'm with the PCPI Foundation, and I have my colleagues here, Yvette Apura and Diedra Gray, and then we also have our clinical expert, Dr. Stephen Persell, who will be providing the overview. Dr. Persell?

DR. PERSELL: Yes, hi, everyone. Sorry I couldn't be there in person. So I was

working with the PCPI's work group that made the original version of this measure, and so I think that the highlights of -- so this is a measure of influenza vaccination rate that's suitable for individual clinicians and office practices. And I guess, what level of detail would you like?

MS. MUNTHALI: Two to three minutes more, sorry.

DR. PERSELL: All right, because I can make it pretty short. So this is basically reporting on patients that are seen in office practice during October through March, and the receipt of influenza vaccination or the documentation of medical, patient, or system reasons for not administering the vaccine.

In the data that looked at this, there's still a large gap. Performance is only about 50 percent, and this is substantiated by comparing it to data that the PCPI compared to BRFSS data which looked very similar in terms of the number of adults and children six months and older getting the vaccination.

And the other thing that's notable is that there are some gaps between groups, with some non-Hispanic, white minorities getting lower rates of vaccination and adults getting lower rates of vaccination compared with children. And I think that's probably enough from me.

MS. CHAVARRIA: Thank you, Dr. Persell.

DR. PERSELL: Did you want more of the technical aspects of the measure, or is that what you were looking for?

MS. CHAVARRIA: No, I think that's fine. I think we wanted to turn it over to the lead discussants now.

MEMBER MOLINE: I have a question, and maybe the discussants -- but it's probably easier. Is this all visits, or is this primary care? Yes, I mean, who -- well, I guess they'll be talking about it with a numerator and denominator, but was it -- that's just a question. When you were developing the measure, was it any visit to a health care provider, or

was it -- that's being marked?

DR. PERSELL: The measure, my understanding is that this could be applied by groups that seek to evaluate the delivery of influenza vaccination among their care delivery systems. So while we expect a large uptake in primary care, this certainly could be applicable to many other sub-specialties and care settings, but the measure's -- and the measure really is based on the presence of preventive care visits or having two E&M visits.

MS. CHAVARRIA: And in the ambulatory center, and then also within home health care and nursing.

CHAIR McINERNY: Can I expand on that a little bit? I think, you know, in many instances, if it's a health maintenance visit, the clinician is liable to ask a patient about immunization status, and if they're not immunized, they would recommend it or give it, but if it's an illness or an injury visit, many times they don't go into that.

Now, we have learned that that's a missed opportunity, and in fact, and particularly for flu because of the problem of the seasonality and the fact it has to be repeated every year, that it really -- they should ask about flu vaccine at least for every visit, but I don't know if this captures that or not.

DR. PERSELL: So if one were to try to perform well on this measure, one would have to really focus on all visits during the window because one never really knows whether someone's going to have two visits.

And so I would say yes, this really strongly encourages clinicians to address influenza during the active season, and it also accounts for the fact that someone could deliver good care, but there's medical exceptions or patient exceptions, namely patients not willing to receive the vaccine, that can be measured and then tracked, and I believe the rates of those exceptions were quite small, only about three percent recorded that.

But yes, it gives you a way to track when it's not clinically appropriate or acceptable to a patient to deliver a flu vaccine, but it does really promote it at all visits, even though not every single person that makes one visit will qualify for the -- at least the electronic version of the measure.

MS. MUNTHALI: Steve and Katie, any other thoughts on evidence that you'd like to raise for the group?

MEMBER TEUTSCH: I think there are a couple of things. One is the point that was just made, that they get triaged by the different reasons is important because, you know, it's important to understand whether there are patient reasons for opting out that need to be dealt with.

And I too have the same question about the specialty and the locale of the individual provider because there is likely to be very vast differences between a primary care provider and others, and I didn't see really any breakout or

discussion of that issue, and as Tom said, those are all missed opportunities.

But it looks like it's both at an individual clinical clinician level as well as at a facility or practice level, so, you know, it would be interesting to see some of those kind of breakouts.

MS. MUNTHALI: Matt?

MEMBER STIEFEL: It's still just a question about the specification of the measure. So does it count if a person comes in twice during that period and is documented in the electronic medical record of having had an immunization?

DR. PERSELL: I believe if you record that a person had an immunization during that season, that that would satisfy --

MEMBER STIEFEL: Yes, it's -- oh, I see.

MS. CHAVARRIA: Yes, that's right.

MEMBER STIEFEL: Okay.

DR. PERSELL: It also promotes --

helps the clinicians to ask and record vaccines received elsewhere, which is a big problem with flu vaccine measurement, which is that so many patients receive it at work places and places in the community.

MEMBER TEUTSCH: I think, Matt, your point is also, if I hear you right, is also an important point, because there is a big difference between getting your vaccination in September or October, and coming in and getting it in March. That's -- you know, you've missed most of the season.

So, you know, I think there are some interesting issues regarding timing, although I don't know how to incorporate them into the measure themselves, but if you get it on your second or third visit when you should have got it on your first, that's a problem too.

MS. MUNTHALI: Diedra or anyone else from the development team?

MS. APURA: I just wanted to remind everyone they are reviewing the registry version

of the measure.

CHAIR McINERNY: One other issue on the pediatric side is that for the first time around for influenza immunization, that is between, roughly between six months and two years of age, to be adequately immunized, you need two immunizations with influenza. One is not sufficient. I don't know if your data collects that or not.

DR. PERSELL: As currently written, it does not. It does not have a separate criteria for infants and up to two.

MS. MUNTHALI: Any other comments or questions on evidence, just on evidence for now? Okay, so perhaps we do take a vote. There are a number of clarifying questions. This is a maintenance measure, but we will vote on evidence for 0041. One is high, two is moderate, three is low, and four is insufficient. Voting is open, and it looks like 13.

For Measure 0041, 2 voted high, and 11 voted moderate on evidence, so this passes

evidence. We'll now discuss performance gap, and I'll turn it over to Steve and Katie.

MEMBER TEUTSCH: We already were presented some of the information on performance gaps, and there are some modest performance gaps among all of these different demographic groups, but the biggest thing is there's a big gap between all of them and what needs to be.

So there's a big performance gap, less so among, you know, a modest amount for the disparities, and so you get a hint of some of the same anomalies that we discussed under immunization generally.

MEMBER SELLERS: I would just add there was a fairly large gap between states too. I saw 39 percent in Florida and 59 percent in South Dakota, so there's a big geographic gap.

MS. MUNTHALI: Other comments or questions on gap? Okay, I think we're ready for a vote. One is high, two is moderate, three is low, and four is insufficient, and we're looking for 14 votes.

So for performance gap for Measure 0041, 11 voted high and three voted moderate, so we move onto reliability. Steve and Katie?

MEMBER TEUTSCH: So, I mean, it's much the same as we talked about that. I was a little perplexed here because they used the same information on reliability as they got out of the ESRD. That's what's stated here, and it struck me as a little odd, not that they aren't similar issues, but that it wasn't immediately obvious that that was as relevant. I don't have any particular reason to doubt that you can get this reliably for the same reasons.

MS. GRAY: Hi, I just wanted to clarify that the original data that was submitted for testing was from, I believe, 2012. That's the one that included the ESRD data. At the time, we were performing testing using the inter-rater reliability method, which is a lot more expensive.

It requires a lot more resources. You have two manual abstractors visiting a site and

abstracting the data. So we were trying to multipurpose our data sets, if you will, at the time. We have since submitted some updated testing information that's signal-to-noise ratio analysis, and that was based on 2014 data. The SNR results from the 2014 data were -- okay, so--

MEMBER TEUTSCH: You're reporting 80 percent reliability when the minimum level of quality reporting advanced 0.99 evaluated with the average number.

MS. GRAY: Yes, so the reliability was still high when we performed the signal-to-noise ratio analysis. I think part of the confusion was with the ESRD population being included in the original data, and the fact that the -- I'm not sure if it's always clear when we add in the updated data with the old data. I think sometimes it gets confusing.

MS. MUNTHALI: Any other questions about the specifications? Oh, Arjun?

MEMBER VENKATESH: So does that mean the safe way to evaluate this is just look at the

registry data that is specific to this type of care setting that had signal-to-noise ratio analysis? You can interpret that 0.8 as high and just kind of ignore the kidney disease stuff?

MS. GRAY: I'm sorry. Okay, I would need a little bit more clarification on your question.

MEMBER VENKATESH: So we're asked to, you know, assess the reliability of this measure. Can we just assess the reliability of the measure based on the data you have from the registry data, which is, I'm assuming, not the ESRD but some sort of office practice registry data, and just ignore the kidney disease stuff from before?

MS. GRAY: Yes, sorry. Yes, that was a previous testing project from years ago. So the data that we submitted from 2014 is from the PQRS reporting program.

MS. MUNTHALI: Any other questions, comments? So I think we're ready for a vote on reliability. One high, two moderate, three low, and four insufficient. I'm looking for 14 votes.

So for Measure 0041 reliability, six voted high, and eight voted moderate, so it passes reliability, and we'll move onto our discussion on validity, and I'll ask Steve and Katie to lead us in that discussion.

MEMBER SELLERS: Sure, okay, now my computer is acting up. No, I've got it. Okay, so it's a maintenance measure with new validity testing provided. The specifications align with the evidence. It was tested at the measure score level, and they did face validity only.

The face validity was assessed by a nine-member expert panel from the PCPI Measurement Advisory Committee. Committee members were asked to rate their agreement with the following statement: "The scores obtained from the measure as specified will provide an accurate reflection of quality, and can be used to distinguish good and poor quality." It was a five-point Likert scale from strongly disagree to strongly agree.

The results of that, of the nine

members of the panel, eight members agreed or strongly agreed, and one indicated disagree. I guess as I was reading this, I was just wondering is there any information about the one dissenter, why, you know, what the rationale was for disagreeing. Do you have access to that information?

MS. GRAY: I don't, unfortunately. We just asked them to complete the rating using the Likert scale, and sometimes they choose to provide information. In this instance, they did not.

MEMBER SELLERS: Okay. There is no risk adjustment. As far as exclusions go, documentation of medical reasons, patient reasons, and system reasons for not receiving the immunization are in the exclusions. As the developer mentioned, the exclusions were about three percent, but I was wondering about the system reasons. An example of that was the vaccine not being available. Are there other system reasons that can be listed?

DR. PERSELL: So my understanding is that the element that's reported is, "Not done, system reason," and that the absence of the vaccine not being available is an example, but the exception criteria is just simply, "Not done, system reason",

MS. APURA: Other examples of system reasons are, "not entitled to benefits", "drug not available", and other reasons.

MEMBER SELLERS: Okay.

MEMBER TEUTSCH: Not entitled to benefits? That's almost unheard of these days.

MS. APURA: Yes, this, you know, is one of the examples, and there are other -- I can say, "other", here, "patient on waiting list", but, you know, the doctors -- because this one is like, allows clinical judgment, so the doctor can just document reasons that would fall under that bracket.

MS. GRAY: So just to add additional clarification, the exceptions, like Yvette said, are to allow for clinician judgment, but also the

system reasons really are to ensure that because this is a provider-level measure, that the provider is not penalized for, you know, some larger reason, that the vaccine is not able to be given to the patient.

MEMBER SELLERS: And I guess an important distinction to make here between this and the other flu measures is that patient refusal is an exclusion.

MS. MUNTHALI: Other comments, questions? Barry?

MEMBER HARRIS: And hopefully we're getting ready to vote, but when we get ready to vote, are we going to have the same issue?

MS. MUNTHALI: No.

MEMBER HARRIS: Okay.

MS. MUNTHALI: We changed the number, so what Barry is referring to is, because the highest rating we can give face validity is moderate, so we only included the options you have there. So in the past when we did it the first time and it didn't work, moderate was two,

but now it's one. So for moderate, you would vote one, low two, insufficient three.

So 13 voted moderate, and 1 voted low, so for Measure 0041, it passes on validity, and we'll move onto feasibility. Katie and Steve?

MEMBER TEUTSCH: And I think we've discussed feasibility before, although it's not altogether clear to me that these exceptions are necessarily in the record, but if they are, I mean, then it seems pretty straightforward.

MS. MUNTHALI: Okay, it looks like we're ready for a vote on feasibility for 0041. One high, two moderate, three low, and four insufficient.

So 10 voted high, and 4 voted moderate for feasibility for 0041, so we proceed to usability and use. Steve and Katie?

MEMBER TEUTSCH: Oh, this is already in use in the PQRS primarily. They have the ratings, which are around 50 percent, no real issues, so this is fine.

MS. MUNTHALI: So we're ready for a

vote, usability and use: one, high; two, moderate; three, low; and four, insufficient information.

11 voted high and 3 voted moderate for usability and use for Measure 0041, so we'll move onto the final vote, overall suitability for endorsement: one, yes; and two, no. We need two more votes. It's unanimous. 14 voted yes, so this Measure 0041 is recommended for NQF endorsement. Thank you.

So we'll move onto the eMeasure version of 0041. That's measure 3070. Am I correct? Yes, and so for this measure, the evidence base is the same, so what we're going to do is carry over the votes from the claims-based measure, 0041, to 3070, and we'll start discussion on performance gaps.

MEMBER TEUTSCH: Yes, I don't know what to say. The numbers that are presented are basically the same as they were before, so the same gaps exist.

MS. MUNTHALI: Okay, and Steve, you're

by yourself today. John's not with us today, but -- so you're by yourself on this measure.

MEMBER TEUTSCH: I have to channel John? That's not going to be possible.

MS. MUNTHALI: He'll be here tomorrow.

MEMBER TEUTSCH: Okay.

MS. MUNTHALI: Okay, any other comments on performance gap? Any questions for the developer? We do have to have a formal vote on performance gap, so we'll tee that up, so one, high; two, moderate; three, low; four, insufficient.

Okay, 10 voted high and 4 voted moderate for performance gap for measure 3070, so we'll move onto reliability. This is a fully specified eMeasure, so unlike the trial use measure, so the developers have something to say.

MS. CHAVARRIA: Yes, I just wanted to point something out, and this came up with Measure 0041, which is a registry-based measure. In PQRS, it was originally the claims-based measure, and PQRS does not offer the

functionality of taking into account two visits.

But since this measure is an eCQM being used in the Meaningful Use Stage 2, and it is, in fact, also proposed for use in MIPS, the CMS Merit-Based Incentive Payment System for reporting in 2017, this one actually does -- EHRs actually will provide the functionality to take into account two measures.

So the denominator for this measure is just slightly different, and I, of course, made a typo and included the same denominator, but the denominator for this one is actually, and I will read it, "All patients aged six months and older seen for at least two visits, or at least one preventive visit during the measurement period, and seen for a visit between October 1 and March 31," which is similar with 0041. So again, this does provide for two visits.

MEMBER TEUTSCH: So could you explain the numerator? It says, "doesn't include offer and decline." It doesn't talk about the systems problems. It doesn't talk about patient refusals

explicitly. Are those the same?

DR. PERSELL: My take on this is that the numerator criteria is delivery of the vaccine or documented receipt of the vaccine in the current season, and then exceptions that are applied if the numerator is not met would be patient, system, or medical reason, so it's not technically part of the numerator criteria.

MS. MUNTHALI: So, Steve --

MEMBER TEUTSCH: But it is in the electronic medical records that you can distinguish those things?

DR. PERSELL: It requires configuring electronic health records to capture these exceptions.

MS. MUNTHALI: Steve, what you saw at the top of the page was the staff analysis as we were looking at the NQF standard specifications for influenza vaccine, and so we were pointing out where there was misalignment, and so that was part of it. It wasn't the medical reason, or the patient reasons that you included, but just

wanted to note that. Any other questions?

CHAIR McINERNY: Well, I'm a little concerned about just looking at two academic medical centers. We know from other kinds of studies how academic medical centers perform versus how those out in the community -- practices are different.

Sometimes one is better than the other; sometimes the other way around, and I would think it would be better to do the testing, as some of us who practice in communities, some of us call the real world, versus academic medical centers.

MS. GRAY: So, thank you for that comment. I think that there might be a little bit of confusion because the academic medical centers was actually the feasibility testing, and the reliability -- does that say reliability? I can't see that far. Yes, I'm wearing glasses, but I still can't see that. The reliability testing is actually done from a sample from the PQRS program. That's not limited to the two

academic centers.

CHAIR McINERNY: Okay, because it says here under reliability there were two academic centers.

MS. GRAY: Oh, okay, that must be a typo.

MEMBER TEUTSCH: But the reliability that I saw that you're referring to from PQRS, I thought that was the registry data. It looked like at least what we saw here was the same as we saw earlier.

MS. GRAY: It's actually different. PQRS allows for reporting via registry option and reporting separately through an EHR.

MEMBER TEUTSCH: So you're looking just at the EHR portion?

MS. GRAY: Right, so this is just the EHR data, and the reliability results are a little bit different even though they're still -- the reliability is still high for the measure.

MS. MUNTHALI: Other questions, comments? Okay, so we'll proceed with a vote on

reliability for Measure 3070. One is high, two is moderate, three is low, and four is insufficient.

We're looking for one more vote. So we're fine, 13. Someone stepped out. So, 8 voted high, and 5 voted moderate for Measure 3070, reliability, and so now we'll proceed to validity.

MEMBER TEUTSCH: I didn't see much evidence from this, and between this and the other measure that we just passed.

MS. MUNTHALI: Okay, the claims-based? Other questions or comments? Okay, I think we can vote on validity for Measure 3070, high, one -- we're going to read you that because we brought up the wrong slides. This is, again, only eligible highest vote is moderate, so one is moderate, two is low, and three is insufficient because they did face validity.

Okay, 11 voted moderate and 2 low, so 3070 passes for validity, and we're going to move onto feasibility. Steve, any comments?

MEMBER TEUTSCH: It sort of echoes what Tom said a few minutes ago. The testing was done in a single EHR system and in two academic medical centers, so it's, you know, fairly selective, but, you know, there's no specific issues with it.

MS. GRAY: So for feasibility testing, it's a little more difficult to recruit sites to participate in that. We have to identify sites that are not only willing to participate, but sites that have already implemented the measure, plan to implement the measure, and so we recruited the two academic medical centers.

We don't have anything to incentivize their participation, unfortunately, so we have to try and charm them, but we got the two academic medical centers and the EHR vendor. It's supposed to kind of serve as a sample.

And I know ideally we would be able to include, you know, different clinical settings and more, but I will just say that our testing efforts are ongoing, and so it doesn't stop our

recruiting efforts, and our testing doesn't stop at NQF endorsement. We continue to try and reach out and identify participants.

MEMBER TEUTSCH: That raises the question, and maybe it's for NQF than for you, why was this not a testing measure as opposed to a -- one that was -- it looks like it's being presented as one that's ready to go?

MS. MUNTHALI: Yes, I think this is one of those complicated ones where there's a claims-based measure that's already in a program, and this measure was also in a program, Meaningful Use 2, but it had never come to NQF, and so there was a period -- these are the legacy measures that Jason talked about.

We were trying to help the field along while there were requirements out there by the federal government that, you know, they be accompanying electronic clinical measures that accompany the claims-based measures. So we're bringing them back into our process now, and it's a good point.

It's a good question about whether or not this should be a trial use measure, but trial use is only -- it only applies to measures that haven't been implemented. This measure technically has been implemented. It just had never come through NQF before.

So sitting around this table and evaluating it against our major criteria is new. You did do that for the claims-based measure, but not for the eMeasure. So I'm not sure if that answers your question, but it looks like the PCPI Foundation, they also changed their name, so we're trying to --

MS. GRAY: Yes.

MS. MUNTHALI: -- get that right.

MS. GRAY: There was some confusion.

MS. MUNTHALI: So it sounds like you guys are continuing to test in multiple EHRs?

MS. GRAY: Yes, our recruitment and identification of test sites and testing does continue. And I would just like to add that we also included, in addition to the feasibility

assessments from those three participants, we also included the Bonnie testing, which is hopefully helpful to give you an idea of how the measure would perform in a larger environment.

It contains 65 patients, and everything -- we got 100 percent coverage, and all of the patients passed, so hopefully that helps a little bit more and adds more to the feasibility testing.

MEMBER TEUTSCH: I read through one of the Bonnie ones. I think this was the one with all of the patients in all of the different sites, and it struck me as if there's some that's okay to go now, and some that were going to be okay in the future. So I read through it, but I'm not sure it was all that enlightening for somebody like me.

MS. MUNTHALI: So perhaps what we could do, since you guys are in the process of testing, would you be -- do you think you'd be ready to bring forward testing in another EHR by, let's say, your annual update, so like a year

from now?

MS. GRAY: Of a different EHR vendor or a different type of setting?

MS. MUNTHALI: A different -- well --

MS. GRAY: It's a trick question.

MS. MUNTHALI: Yes, it is a trick question. I would rather a different EHR vendor, but what can you do in a year?

MS. GRAY: We can attempt both of those --

MS. MUNTHALI: Okay.

MS. GRAY: -- in a year, and hopefully we'll be able to charm some more people to participate in our feasibility testing.

CHAIR McINERNY: Could you scroll up to see where it was tested again, please? Well, you know, when you look at that, a 619 multi-specialty academic medical centers serving 33 counties, that's quite a few, and the other academic medical center handles over two million outpatient visits and 40,000 hospital stays.

So even though they are academic

medical centers, they clearly have outpatient locations in the real world, so that makes me feel a little bit more comfortable with where you had tested the measures.

MS. GRAY: Yes, the second entity did or does have over 150 clinics and extensive home care operations, so --

CHAIR McINERNY: Thanks.

MS. GRAY: Thank you.

MS. MUNTHALI: Any other comments, questions? Okay, so we'll move forward on a vote on feasibility. One is high, two is moderate, three is low, and four is insufficient. On feasibility, 2 voted high, 10 voted moderate, and one voted low, so for measure 3070, this measure passes on feasibility, so we'll move onto usability and use.

MEMBER TEUTSCH: So usability is for the same, all intents and purposes, similar PQRS. It, you know, part of Meaningful Use Stage 2, so not much different.

MS. MUNTHALI: It looks like we're

ready for a vote. One, high; two, moderate; three, low; and four, insufficient information. So we're looking for two more votes. So for Measure 3070 usability and use, 3 voted high; 11 voted moderate, so we'll move onto an overall vote for endorsement suitability: one, yes; and two, no. We're looking for two more votes.

CHAIR McINERNY: Unlike Chicago, these things are coded so that your vote gets recorded only once even if you vote four or five times, so it's okay.

MS. MUNTHALI: So it is unanimous, so Measure 3070 is recommended for NQF endorsement. Thank you.

MS. GRAY: We flew in from Chicago, so we understand that you're saying.

MS. CHAVARRIA: Thank you, Dr. Persell.

DR. PERSELL: Thanks, the weather is very nice in Chicago today.

MS. MUNTHALI: So we're making progress. I think we're almost caught up,

almost, not quite. So the next measures are also influenza vaccination measures. The first is 0680, percent of residents or patients who were assessed and appropriately given the seasonal influenza vaccine, short stay.

The steward is CMS and the developers are RTI, and they'll be doing the long stay measure, 0681, soon afterward. So if you could give us a two to three-minute intro to your measure, the first one?

DR. BYRNE: I'm here with my colleagues, Amy Helburn and Laura Smith. We're with RTI International, measure stewards for CMS. The cross setting measure, NQF 0680, reports the percentage of short stay residents or patients who were in the facility for at least one day during the most recently completed influenza vaccination season, I'll refer to as the IVS, and who were assessed and appropriately given the seasonal influenza vaccine.

The IVS is defined as beginning October 1 or when the vaccine first becomes

available, and ends on March 31 of the following year. The measure is the aggregate of three separately calculate sub-measures to reflect the process by which a patient or resident is assessed and appropriately given the influenza vaccine.

The three sub-measures are residents or patients who received the vaccine either in the facility, hospital, or outside the facility or hospital, patients or residents who were offered and declined the vaccine, and residents or patients who are ineligible to receive the vaccine due to contraindications.

The quality measure 0680 was endorsed for use in the nursing home setting in 2011, and then was expanded for use in the IRF and LTCH settings in 2012. This quality measure is based on the NQF's national voluntary standards for influenza and pneumococcal immunizations.

Influenza is associated with increased morbidity and mortality in high-risk adult populations, people with comorbidities, and the

elderly. Annual seasonal vaccination is an essential element of a multi-faceted approach for preventing the spread of influenza, and an effective preventive measure against influenza-related hospitalization and death.

Public comment and subject matter expert input received on this measure was predominantly supportive of continued endorsement of this quality measure because it improves the quality of care to patients, is not burdensome to implement, and retirement of this measure may result in fewer residents and patients being vaccinated for influenza. The measure is also feasible to implement, with only minor or very rare unintended consequences.

The quality measure is based on assessment of nursing home patients, inpatient rehabilitation facility or IRF patients, and long-term care hospital or LTCH patients using standardized influenza items.

The influenza data elements used for this quality measure are the same across the

instruments, and have been shown to have high reliability and high validity.

The denominator consists of the patients or short stay residents, 100 days of age or older, who are in the facility for more than one day during the IVS, and the measure is based on episodes for short stay residents with 100 or fewer days of nursing home care, and for stays of all lengths for LTCH and IRF patients.

The quality measure scores for the percent of residents or patients assessed and appropriately given the vaccine for the 2014-2015 IVS was 91 percent for IRF, 74 percent for LTCHs, and 81 percent for nursing homes for short stay patients. A very small percentage of residents and patients received the influenza vaccine in the facility, less than nine percent across any of the three settings.

About one-quarter of the IRF patients and short stay residents declined the vaccine, and in LTCHs, about 15 percent of patients declined the vaccine. A very small proportion of

patients and residents did not receive the vaccine due to medical contraindications.

In testing reliability and validity, the results demonstrated acceptable to high reliability and validity of both the data element and the quality measure across each setting. For all three settings, two-thirds or more of facilities had scores that differed from the national mean.

We'd like to point out to the committee that we did provide results of testing for the items, or the two influenza items, validity and reliability testing, and have sets of kappa scores for both of the items. Kappa scores were high for the reliability and validity results, and this is based on the testing of the MDS 3.0.

For the 2014-2015 IVS, the percent of facilities with a perfect score, meaning all residents and patients were assessed and where appropriate vaccinated, were low for nursing homes and LTCHs, and for IRFs were around 13

percent. The between facilities' differences in the QM scores were found to have a small to medium and significant effect on QM scores across the setting.

There was a moderate and statistically significant correlation between the short stay and long stay influenza measure for nursing homes.

There is opportunity for improvement of this measure by assessing and vaccinating more patients and residents, and reducing the percent of those who decline. We found that 10 percent of IRFs had more than 34 percent of their patients decline the vaccine, and 10 percent of nursing homes had more than 42 percent of their short stay residents decline the vaccine.

Disparities in nursing home residents' vaccination status were observed over 10 years ago, and there is continued evidence of disparities in whether post-acute residents and patients are assessed and receive the vaccine.

Males, whites, and older individuals

were more likely to receive the vaccine, and women, persons of black race, and Hispanic ethnicity, and younger individuals were more likely to decline the vaccine across all of the settings.

Further, we did find across the settings that facility characteristics associated with the higher performance, that is in the top 10 percent of patients and residents receiving the vaccine, or in the lowest 10 percent of patients and residents declining the vaccine, were found to be smaller sized facilities, more likely to be nonprofit or government ownership, and in rural locations. Thank you.

MS. MUNTHALI: Thank you. Marcel, do you want to start the conversation on evidence?

MEMBER SALIVE: Okay, thanks. Those last parts, I think, clarify this first section on performance gap. We're not discussing the evidence, right? So --

MS. MUNTHALI: Yes, we're saying that for the record. The evidence will carry over.

MEMBER SALIVE: Yes, so the performance gap, I think she just mentioned those figures, and I felt they were very constructive, that there are wide differences amongst the facilities in the percent vaccinated, and then there are disparities evident. So, to me, that answers our questions.

MS. MUNTHALI: Patricia?

MEMBER McKANE: I agree.

MS. MUNTHALI: Okay, so it looks like there are no comments or questions, so we can vote on performance gap, except for Matt.

MEMBER STIEFEL: But just this has the same source of confusion for me, that it includes in the numerator those people who declined, so it would just be, I guess, the same clarification for this measure.

MS. MUNTHALI: So adding the computed and reported to it as per the NQF specifications, that would help to clarify the confusion.

MEMBER SALIVE: So they have composite measures though, so it can be teased out in this

case. I thought that it was reasonable.

MEMBER McKANE: I did as well, and I thought -- I appreciated the tables with the analysis that was done showing the percent that refused and all the exclusion factors because if you don't count them, then you're assuming they're like the rest of the population, so if you do show it, then you're --

It's the difference between a statistician and an epidemiologist. You know, statisticians are going to love this, so, epis, maybe not so much, but I thought it was -- I appreciated that information in the tables.

MEMBER SALIVE: Also, I think you could, you know, take the data, if you're running that institution, and focus your QI on what to look at. So if you have a high amount of refusals, focus on that. If you have a high amount of contraindications, that's probably not really correct, so you could focus on that. So I mean, I think it seemed quite reasonable.

MS. MUNTHALI: Any other comments,

questions? Okay, so we will vote for Measure 0680, one, high; two, moderate; three, low; four, insufficient. So for performance gap for Measure 0680, 11 voted high and 3 voted moderate, so we'll move onto reliability, and I'll turn it over to Marcel and Patricia.

MEMBER SALIVE: They had, I think, quite a lot of data for the reliability testing, and, you know, it was in the order of many millions of people. So I, you know, it's hard to find a complaint, I think, with that. I was confused a little bit, I think, on the validity testing, but we'll get to that next.

MEMBER McKANE: I just, you know, wanted to point out there was a question from the NQF staff, and I just was hoping that somebody could clarify it for me. There was a difference in how the numerator, I believe the numerator, the difference in the specifications between the different hospital types, that the nursing home, I believe they only counted the most recent visit, versus the long-term care hospital and

inpatient counted all of the visits.

So I was wondering about the difference in how -- it doesn't appear that one was duplicated and one was -- and what was the rationale, a vast difference in data sets, or what the rationale is for that?

DR. SMITH: Hi, this is Laura Smith. That's a great question. There is kind of a mixture of reasons, but the primary reason is that the short stay nursing home measure harmonizes in terms of the episode definition with other currently publicly reported nursing home measures on the Nursing Home Compare site, which does use that episode definition and only the most recent.

And during the development of the IRF and LTCH measures which are much more recent, within, I guess, was it 2012, that -- the nursing home measures are basically built off of MDS 2.0 measures. They're sort of -- we have this, the weight of history, and also there are particular reasons why it's advantageous for the survey

process for a nursing home to use that episode snapshot.

And so the rationale for the IRF and LTCH development was slightly different in terms of thinking about a post-acute care, very specifically post-acute care measures that would be trying to say, "Okay, every time you have an opportunity to do the right thing, have you done it?" and then also looking at this very specific admission to discharge period.

MS. MUNTHALI: So, just wanted to further that discussion. The staff's concern was that we require that measures be specified, they be tested on how they're specified, so all of the settings of care, that they would be specified.

We saw that for nursing homes, you did use MDS 3.0 data to test, but there were only two data elements used there as well. Can you talk about your plans for further testing for IRFs and LTCHs?

DR. SMITH: Okay, and so that's a separate question than the episode question.

Okay, so we have included -- we did include measure-level testing for all three data sets for reliability and validity. That was what we primarily focused on.

We also cite data that shows that there is significant overlap in the populations that receive services across long-term care at hospitals, inpatient rehab facilities, and skilled nursing facilities, and so to justify the use of the MDS data.

I'm trying to think. Colene, was there anything else that we should add on that discussion?

DR. BYRNE: That's a similar population, although I don't believe we should be applying the testing from the MDS to the --

MS. MUNTHALI: So in terms of the data elements, are you using all of the critical data elements? It's just the two that we saw in there, so we just want to make sure we're not missing anything. So are you saying that they are basically generalizable from MDS to IRF to

long-term care?

DR. SMITH: Yes, they are identical -- so they are identical items, so the item-level testing is not dependent on that episode definition versus stay definition. That's sort of independent because the item-level testing is within a single assessment, and the items are the same across the different settings. There was something else that you had just asked.

MS. JOHNSON: We had a question too about your score-level testing that you did. Most of today we have seen kind of the Adams signal-to-noise methodology. You guys have done something a little different for your score-level testing. So I think the question that I would have is just real quickly why did you decide to do it the way you did it?

So you were basically testing and looking at how many groups would be significantly above or below the mean, not exactly quite getting to, "Can you differentiate providers?" It's a little bit different question than what

you're answering with that analysis, so can you connect the dots for us on that one?

DR. SMITH: Sure, so I can't remember the paper exactly, but there -- it may have been a Zaslavsky paper that offered sort of different alternate ways to examine reliability, and one way sort of thinking about reliability is telling you when you're thinking about it, at least at the performance measure level, sort of how much signal is there relative to noise, and so by actually calculating the confidence interval for every single provider, you are actually basically depicting the amount of uncertainty you have around each of those measure scores.

And so if you see that once you've actually said, "Okay, the range of confidence around these scores looks like this and relative to the mean," if you're seeing that a lot of them don't overlap that national mean, then it suggests that there is -- there are differences that could be attributable to the characteristics of the provider rather than sort of random noise.

It's not as much of kind of where you -- where people tend to go with it, also given that it's somewhat of an intensive way of doing the analyses, but it is something that actually has been recommended as a potential way of examining the reliability of your measures.

We did include the eta statistics as well in recognition that different people do things different.

MS. JOHNSON: Yes, and Patricia made me laugh with the battle of the statisticians and epidemiologists, but, yes, the eta statistic is actually new to me, so I was just a little bit curious about you did one-way ANOVA. I couldn't tell from your description did you do -- is it a random effects ANOVA? Did you nest within facilities? And is there a problem with having different numbers and patients within each facility when you're doing that kind of analysis?

DR. SMITH: That is a good question, and -- yes, so I'm going to throw that to Dan Barch who is on the line.

MR. BARCH: Hi, thank you, Laura. It's a random effects, and, well, it's not technically an ANOVA, for just that reason, that we don't have equal N in every group.

But it is a generalized linear model, and I believe that our effect is robust to the different assumptions of the ANOVA. And so the fact that we found a significant effect would hold up if we did have equal N and normal distributions.

MS. JOHNSON: Thank you, Dan. Your sound is a little bit low in here. Can you repeat that very first sentence that you had. You said it was or was not the random effects ANOVA?

MR. BARCH: Okay, sure. Is this better?

MS. JOHNSON: Not much, but I'm cocking my ear, I'm trying to pay attention.

MR. BARCH: Okay. Yes, it would be a random -- well, the effect doesn't matter so much in the calculation of the eta-squared. That

would affect the omega statistic but not -- the eta would be calculated the same way, other way. And also that it's a, this would be a generalized linear model and not actually an ANOVA.

MS. JOHNSON: Okay.

MR. BARCH: Because you don't have equal N.

MS. JOHNSON: Okay, so it was a little bit of mistake in terms of your methodology there. You did not do an ANOVA, you did a hierarchical general linear model of these effects?

MR. BARCH: No, no, we -- it is an ANOVA; it's just not, you can't technically call it an ANOVA. But the methodology is, you know, it works the same way with the software. The idea of it is very much an ANOVA. We're looking at within and between group variants.

MS. JOHNSON: Okay.

MR. BARCH: It's not a hierarchal model in the sense that we're predicting anything.

MS. JOHNSON: Okay, got it.

MR. BARCH: It's just that very technically it's not -- it's an ANOVA in all but name.

MS. JOHNSON: Okay. I think what we'll need to get from you is just maybe the paper that you were talking about from Alan, that talks about if he was -- I know Alan has done inter-unit reliability of the f-statistic. This is a little bit different, but --

DR. SMITH: So I can get you the paper for the confidence interval analysis. But that's not the same thing as what -- okay, all right. Just double-checking, because that's not what Dan was talking about.

MS. JOHNSON: Right. So we've gotten into the weeds here, but basically what the developers have done is they have done some score-level testing. So from three different data sets from the actual settings where you have specified the data.

And because it is, your methods are a

little bit different than what we're used to seeing, that's why we're asking them a little bit more detail to try to understand what they've done.

So given, in terms of the first method that they did with the differences in means, with the confidence intervals, what you can say there, since you had some that were statistically greater than the mean based on the confidence interval and some that were statistically lower than the mean, you can say that there's at least some providers that are different from each other, right.

And that's what we're trying to get at with reliability. We're trying to say, Can you differentiate between providers. It doesn't give you maybe quite the same kind of information as the Adams signal-to-noise that we usually see. It gives you some indication.

And then they've done a second method with their eta statistic, with their ANOVA that's not an ANOVA. Again, that is beyond, that's new

to me, so, relying on Dan's description of what that is, that would, I think, suffice.

It would be an appropriate method, because it is looking at variation between, versus total variation, which is another way of saying what we're interested in with reliability. So apologize for getting into the stats weeds here.

MEMBER SALIVE: Since you, you know -- since it's a renewal of a measure that was approved before, I mean, they went, I think, pretty far in my opinion. I, you know, despite their confusion to you.

MS. MUNTHALI: Yes, they did, but they updated it by adding the two additional settings. And that's why we were concerned. We wanted to make sure that they tested appropriately for those settings of care.

MS. JOHNSON: Right, so just to beat it to death.

MS. MUNTHALI: They used different data sources. And so we needed to make sure that

we can understand what happened in terms of the settings, and to see whether or not they were very complimentary, as they've been saying.

But Karen is not really into more of the weeds than we ever get into, so I think we're satisfied.

Questions, comments? Okay, so, I think we can vote on reliability for Measure 0680. One is high, two is moderate, three is low, and four is insufficient. And I think we're looking, yes, 14 votes, and yes.

(Voting.)

Two more votes.

CHAIR McINERNY: Vote again.

(Voting.)

MS. MUNTHALI: Okay, so consensus is not reached on reliability. We have one high, six moderate, five low, and two insufficient. We will continue voting and resolve this during the post-comment call. So we'll go to validity.

MEMBER SALIVE: So validity has, I think, well-specified elements. You know, you

were getting into some of that same discussion, I thought, in the stats. I did not see threats to validity that were of concern to me. Did you have comments?

MEMBER McKANE: No, I thought that the validity was fairly good. I thought that, I'm trying to remember this study, but the exclusion criteria were fine. I don't think I really had any questions or concerns with the validity on this measure.

MS. MUNTHALI: Any other comments?

MEMBER SELLERS: I guess I have a question. Which, you know, when I'm looking at the measure worksheet and I see the staff rating the IRF and the LTCH as insufficient, could you, do you have an update to that preliminary rating, based on the conversation we've already had?

MS. MUNTHALI: She's asking for an update, yes.

DR. SMITH: So we have, I guess we've talked some about how our rationale for using some of the item-level analysis for nursing home.

I think the other area is the face validity, which I think Colene has some talking points to address the face validity question.

DR. BYRNE: Well, referring to public comment and subject matter experts. Yes. So we had almost all of the subject matter experts in public comment were very supportive of continuation of the measure, felt that it was important to prove and maintain quality of care.

But one of the subject matter experts expressed any concern about the measure.

DR. SMITH: And that was based on interviews with seven subject matter experts, with representation across different clinical specialization and settings.

DR. BYRNE: Right. I think we had seven public comments. And then we had spoke with around 13 subject matter experts from all of the settings in different disciplines in qualitative interviews.

MEMBER McKANE: I think we, I mean, I think it's a great measure. I think the concern

is, is this additional, the long-terms care, I'm going to get the acronyms all mixed up. But the two that were at the nursing home, the two additions, you know, is that going to measure what we want it to measure in the validity?

And I think there were some concerns that the staff expressed about how missing data was handled or, you know, about the validity, how validity was tested. And that's kind of what I'd like to hear clarified. And if not today, at least in the future, so I have a little bit better understanding.

Again, my stats knowledge is about this big, so. You know, I know it's important, but I just feel like I need a little bit more information on it. I think that, you know, based on just like intuitively, I feel like this is probably a very valid measure. But just to have the proof. Pardon my epi brain.

MEMBER SALIVE: Well, to me the settings are, you know, there are some differences technically, but these are other, you

know, they have a lot of similarities, the data set that they're filling out for this measure is very similar.

So there's no, you know, other than inexperience or something. I mean, I don't know why you would think they would be different. Like, that's what I'm trying to say. And so you've run some of the same analyses. You know, their payment model means they have to fill this data out, right.

So it's not like, you know, they have, that's part of how they get the data. So it should be valid, as well as the other source, which we already said was okay.

DR. SMITH: I just realized I had one other consensus source to contribute to the face validity, which was the ad hoc review that was done by NQF to expand the measures for IRF and LTCH in 2012. Had seemed slightly tautological, but at the same time, it seems worth mentioning.

MS. MUNTHALI: Arjun?

MEMBER VENKATESH: I agree with Marcel

about the data. Between the two, they are very similar. These are data sets that are, there's plenty of incentives in place to get it right and get all the data. So I'm not as concerned about the long-term care hospital distinction.

I guess what I'm left with I think, here, and I'm just trying to summarize this now, like distill to what I can make a decision on is, there's a correlation between performance on this measure and the pneumococcal measure. And so that is sort of moderate, maybe gets a little bit of construct built to the measure.

There's no assessment of the measure with an outcome. It's fine, that's not true for a lot of measures. The only real thing that would meet face validity testing, it sounds like, is the 13 experts that were interviewed. I don't think public comments should be used for face validity.

I don't think, like you said, I don't think a previous NQF review counts for face validity. But it sounds like you had a 13

expert, tech expert panel that reviewed this and said, Yes this has face validity. Because then by that, we would give this moderate, as we've done for other measures.

MEMBER SALIVE: I think there was a paper which had real outcome measures, and that was, I cited it in my review. It's in there, and it's from like one million people who got vaccinated. And, you know, and at the proportion, each percentage, higher vaccination resulted in lower hospitalizations for pneumococcal and immunization -- and influenza from that population.

To me, that's an outcome measure for this relevant and, you know, it was a published paper out of this same data set. So to me that was strong.

MS. MUNTHALI: Any other comments? Karen?

MS. JOHNSON: Arjun once again I think hit the nail on the head. This when -- the face validity, I think the question that I had in

terms of what you guys did. Let me read you what our guidance is for face validity. Hang on just a second.

Face validity, the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process by identified experts and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

I think it was unclear to us, in terms of what you did for face validity, what did you actually ask your experts. That part is still, I think, a little fuzzy to us.

So again, we want to know if you asked them or in some other way got from them their agreement that the measure is able to distinguish good and poor quality. So that's the face validity piece.

The score level validity that you did is absolutely fine. It is a form of construct validity. But it was only for the nursing home

population.

So then the question is, can you or did you do something similar for the IRF in the long-term care hospitals? We didn't see that, so that's why for the IRFs and LTCHs we had insufficient. So those are the two outstanding questions from the staff point of view.

DR. BYRNE: The subject matter experts were primarily asked about is the measure important, is there value in the measure, is the measure impacting processes of care, is it resulting in the staff assessing and vaccinating the patients or residents, and about unintended consequences burden.

And should the measure, do they recommend that the measure be maintained or retired. And all but one did suggest that the measure be maintained, that it's important that processes are in place, specifically asking about whether it differentiates in terms of care, of providers.

That was not a specific question, I

think in part because it's a process measure and the evidence, as you said about, you know, how fit improves the care of the patients.

There's a general, you know, support for vaccination but not -- we just didn't, you know, ask them if they thought it was good for differentiating quality.

MS. JOHNSON: I did want to address Marcel's point about the paper that you talked about. If that paper did actually look at facility level outcomes, so not at the patient level but looking at facilities.

If facilities, you know, had higher rates of immunization and therefore had better patient outcomes, I think that absolutely would count in what we would look for as floor-level validation as well.

MS. MUNTHALI: Any other questions? Okay, so I think we're ready for a vote on validity for Measure 0680. One is high, two is moderate, three is low, and four is insufficient. Okay.

(Voting.)

So it's 50-50. So one high, six moderate, four low and three insufficient. This as well is consensus not reached.

We will resolve these issues during the post-comment call. And if there's anything the developer can do to clarify their submission, you can do that during this period, during the comment period. And the committee can discuss that. And we'll re-vote on the reliability and validity, and then have a final vote.

So we'll move on to feasibility.

MEMBER SALIVE: So I think we talked about the element, and it's widely used. So it's very feasible.

MS. MUNTHALI: Pat, you agree? Okay, so let's move forward and vote on feasibility for Measure 0680. One is high, two is moderate, three is low, and four is insufficient.

(Voting.)

Thirteen voted high and one moderate for feasibility for Measure 0680. So we'll move

on to usability and use.

MEMBER SALIVE: So as I said, it's in the Nursing Home Compare website, and is used there, and also for payment in certain settings.

MS. MUNTHALI: Okay, I don't think there are any comments, so we'll vote. Usability and use for Measure 0680, one high, two moderate, three low, and four insufficient information.

(Voting.)

Two more votes, and one more. So 12 voted high and two voted moderate.

So because we did not reach consensus on two major criterion, we won't take an overall vote. And we'll work with you on, you know, what revisions you can make by the time the post-comment call comes around.

So we'll move right into Measure 0681. It's Percent of Residents Assessed and Appropriately Given the Seasonal Influenza Vaccine, Long Stay. Also stewarded by CMS and developed by RTISA.

DR. HELBURN: Thank you. The Long

Stay Nursing Home Process Quality Measure, NQF 0681, reports a percentage of long stay residents who are in the facility for at least one day during the most recently completed influenza vaccination season, who are assessed and appropriately given the seasonal influenza vaccine.

As noted, the IVS begins on October 1 and ends on March 31 of the following year. The measure is the aggregate of three separately calculated sub-measures, which are the same as described for NQF number 0680.

The evidence of importance of the quality measure is consistent with the Short Stay Cross Setting Quality Measure 0680, and is the same as was already described. Public comment and subject matter expert input was predominantly supportive of continued endorsement of this quality measure.

The denominator consists of nursing home long stay residents 180 days of age or older, who have had 101 or more days of nursing

home care during the IVS. The national mean facility level score on this quality measure was 91.5 percent in the 2013-2014 IVS, and 93 percent in the 2014-2015 IVS.

For the 2014-2015 IVS, the percent of nursing homes with a perfect score was 20 percent, and 66 percent of residents received the influenza vaccine in the facility, 14 percent declined the vaccine, 11 percent cited having received the vaccine outside of the facility, and less than 1 percent did not receive the vaccine due to contraindications.

Testing results demonstrated acceptable to high reliability and validity of both the data element and the quality measure. Sixty-one percent of nursing homes have scores that differed from the national mean in the 2014-2015 IVS.

Between facilities, differences in quality measure scores were found to have a medium to large and significant effect on quality measure scores.

There was a moderate and statistically significant correlation between the long stay and short stay influenza vaccine measures for nursing homes.

As noted, there is evidence of disparities in the overall quality measure, and whether residents receive or decline the vaccine. A small but statistically significant difference was found in the likelihood of being in the numerator by race, Hispanic ethnicity, and age.

White and non-Hispanic residents were found to be more likely to be in the numerator. White and older individuals were more likely to receive the vaccine, while black and Hispanic individuals and younger individuals were more likely to decline the vaccine.

Opportunities for improvement with this quality measure may be small. The quality measure score has been between 91-93 percent over the last four IVS. In the 2014-2015 IVS, the quality measure score at the tenth percentile was 83 percent, and at the 90th percentile was 100

percent.

However, as observed through the disparities analysis, there are opportunities for improving vaccination rates among black and Hispanic individuals by reducing rates of decline.

And there is further room for improving rates of declining the vaccine in that 10 percent of facilities, around 1400, have more than 26 percent of their residents decline the vaccine, as compared to 10 percent of facilities that have 0-7 percent decline.

Furthermore, analyses of facility characteristics among the lowest and highest performing facilities indicate important and significant facility characteristic differences, with the lowest rates of decline among residents of smaller, nonprofit, and government facilities, and the highest rates among larger, for-profit facilities.

Rates of decline were also lower among residents of facilities in rural locations.

MS. MUNTHALI: Thank you. So I'm assuming that we're going to skip evidence and accept the prior evidence, and move on to performance gap. Marcel.

MEMBER SALIVE: Yes, so I agree that maybe it's a small difference, and 20 percent I guess had about 100 percent rates. But yes, that means 80 percent didn't. So I think there's still some gap. And I believe also the SES and racial disparities were in evidence, as was stated.

MS. MUNTHALI: Patricia.

MEMBER McKANE: I agree as well. And plus, I think also it's important just because it does seem to be a measure that maybe have a smaller performance gap. It's still important to measure this and to monitor.

MS. MUNTHALI: Other comments, questions for the developer? Okay, it looks like we can move on to a vote on performance gaps for Measure 0681. One is high, two is moderate, three is low, and four is insufficient.

And we're looking for 14 votes. I think Sheila, you're still setting it up, right? Are you? Oh, you haven't gotten to it. So just a minute.

CHAIR McINERNY: While we're going to the slide, you know, noncompliance is sort of the flip side of clinician education of patients.

And I wonder if, when you're reporting back to nursing homes, those that have higher rates of refusal, does somebody figure out or somebody point out that there are ways to reduce those rates of refusal by working more with the patients and explaining to the patients better the need.

Perhaps use some motivational interviewing, those kinds of things that might give you a better compliance rate.

MS. MUNTHALI: Thank you. The slide is up now. One is high, two is moderate, three is low, and four is insufficient.

(Voting.)

We're looking for one more vote.

Okay, we got it.

One high and 13 moderate for performance gap for Measure 0681, so we'll move on to reliability. Marcel and Patricia.

MEMBER SALIVE: So I think it was flagged the same way, but I think there is one slight difference here, which is that this is the long stay measure, and so really people are there for the duration.

So I don't think we have any issue of like leaving and coming back or things like -- I mean it does happen, but this is not that group.

So I had no concerns. I think the missing data was low and there was, you know, and as was said, you could focus on some of these components again.

MEMBER McKANE: I guess I was wondering why, and I don't know if it's something to do with the algorithm that we're using, why this came out as insufficient. Is it because the algorithm doesn't take into account this type of analysis that was done?

Because if we, I mean, you might want to read through this. I think it looks, it sounds good. But then when I go through the actual algorithm, it comes out as insufficient. So just wondering if your staff could comment on that and help me with that.

MS. JOHNSON: Yes, well, so one thing to think about when you look at a staff rating of insufficient, that's not always the horrible thing. It might just mean that we didn't have enough that we felt comfortable to be able to mark one or the others.

So in terms of their score-level testing, they did the same testing score level as they did with the last measure. So if you were happy with that, then you would be happy with the methodology here. Yes, exactly.

In terms of the data element testing, I think the analyst didn't mention -- did you do data element testing for this one as well? Or is it the same?

DR. SMITH: Yes, it's the same

testing, because it's the item-level testing is independent of whether or not it's a short stay or long stay.

MS. JOHNSON: Okay.

DR. SMITH: Measure. And so we, so you'll recall Colene reported that there was inter-rater reliability that was done in the development of the MDS 3.0.

For both items, they are used to calculate the measure, and the kappa statistics are like nearly perfect when you look at a gold standard to a gold standard nurse. And then something like .8 when you have a gold standard nurse compared to a staff.

So we do have definitely have item-level testing for every single item.

MEMBER SALIVE: And this doesn't have these other two settings where we had that concern last time. So I think it is much stronger.

MS. JOHNSON: Right. Just one more question for you. How many data elements are

actually used in this measure? Did your kappa statistics -- you had, what, a couple different kappa statistics, or you had one for numerator, or --

DR. SMITH: We have two items that are used for the calculation. We do exclude people based on age, but that birthdate is coming, I think it's coming from like another source than -- I don't think we did testing.

Yes, we didn't do testing on the birthdate, if we're going to be really precise about this. But we do have, that gets validated in the submission process for the MDS assessment. So the birthdate should be pretty correct.

And then there are two items used for the calculation in this measure. And we did supply four kappas, because of those two different ways that they did the pairings, where you had two trained gold standard nurses and then you had a gold standard nurse and a staff member who actually worked in the participating facility.

MS. JOHNSON: Is that clear to everybody? Anybody have any other questions about? Okay.

MS. MUNTHALI: So we're ready to vote on reliability for Measure 0681. One is high, two is moderate, and three is low, four is insufficient.

(Voting.)

So reliability for Measure 0681, one person voted high, nine voted moderate, two low and two insufficient.

So this measure passes on reliability. And so we can continue our discussion with validity. Marcel and Patricia.

MEMBER SALIVE: So there were minimal threats to validity. The testing was on over two million people. And the paper I mentioned earlier is at the facility level, and so I think this validity is strong.

MEMBER McKANE: I agree.

MS. MUNTHALI: Okay. Any other comments, questions for the developer? Okay, we

can move forward with a vote on validity for 0681. Tom?

CHAIR McINERNY: Validity to only have three, moderate, low, and -- right?

MS. MUNTHALI: They didn't here.

MS. JOHNSON: Yes, the correlation analysis is at the score level, so high would be an option.

MS. MUNTHALI: Yes, it was beyond face validity.

CHAIR McINERNY: Oh, okay. So we can do this way. Okay, I get confused.

MS. MUNTHALI: One high, it's okay. One high, two moderate, three low, four insufficient.

(Voting.)

So one voted high and 13 voted moderate, so this measure, 0681, passes on validity. And we can continue with feasibility. Marcel and Patricia.

MEMBER SALIVE: So yes, they generate this data in the process of care in the nursing

home and it's highly feasible.

MS. MUNTHALI: Any -- well, I guess we're voting. Feasibility, one high, two moderate, three low, and four insufficient.

(Voting.)

So for Measure 0681, feasibility, 12 voted high and two voted moderate. So it passes on feasibility, and we'll continue with usability and use.

MEMBER SALIVE: So again, this is in the Nursing Home Compare website and widely used. There was one comment in the report about some people did not like being asked a lot about do they want the flu shot.

But that is more of a preference and I would say not a harm. So there is no, you know, untoward happenings from this measure.

MEMBER McKANE: Right, and I would just add that I think that it should be continued. There was a question about whether, given its high performance for several years, should it be used to further the goal of high

quality, efficient health care, and I would say, Yes it should.

MS. MUNTHALI: Thank you. Usability and use for Measure 0681. One high, two moderate, three low, four insufficient.

(Voting.)

We're looking for one more vote. Okay, for 0681, 11 voted high and thee voted moderate, so it passed as usability and use.

So now we'll take an overall suitability for NQF endorsement. One yes and two no.

(Voting.)

So for 0681, the committee, by 13 to 1, has recommended the measure for NQF endorsement. So thank you RTI, and we'll be on touch with regard to post-comment.

So we have four more measures for the day. We have scheduled a break. We're going to ask if we can cut that break to five minutes so we can get in votes. Should we push forward, or would you like maybe a ten minute break at most?

CHAIR McINERNY: Break please.

MS. MUNTHALI: So a ten minute break. So we come back at 3:30.

CHAIR McINERNY: 3:35.

MS. MUNTHALI: 3:35, yes.

(Whereupon, the above-entitled matter went off the record at 3:21 p.m. and resumed at 3:35 p.m.)

CHAIR McINERNY: All right, thanks, everyone, for a short break. And, believe it or not, we are on our last full measure. Yes, 1659. Right?

MS. MUNTHALI: Yes, that's it.

CHAIR McINERNY: Influenza Immunization for Inpatients.

MR. DICKERSON: Good afternoon, and thank you, everybody. Appreciate the opportunity to discuss and review this very important measure. NQF 1659 is a CMS national quality measure in the Hospital Inpatient Quality Reporting Program.

It measures the performance rate of

eligibility screening for the seasonal flu vaccine and administration of the vaccine, if indicated, for patients aged six months and older who are discharged from acute care hospital stay from October 1 through March 31 of every year.

This was originally specified for inpatients with pneumonia. It was re-specified in 2011 as a global hospital measure. Now, there are many opportunities for screening immunization within the continuum of health care provider patient interaction.

This measure is the only one that addresses this opportunity in the acute care hospital setting, where patients may be at greater risk than in some other environments.

Of note, the CDC recommends offering the influenza vaccine during hospitalizations to avoid missed opportunities. Now, while performance in this measure has increased over time, there are still disparities in some populations of hospitalized patients.

This measure is important providing

facility-level feedback and maintaining the ability to monitor the performance of the health care system, the prevention of influenza, and complications associated with influenza.

MS. MUNTHALI: Great, thank you. So our lead discussants are Barry-Lewis and Jason Spangler. So, Jason, Barry-Lewis. And I know he's one person.

MEMBER SPANGLER: Okay, I'm going to start, and then Barry-Lewis can chime in as well. We're skipping evidence, right, again?

MS. MUNTHALI: Yes, that's the motion.

MEMBER SPANGLER: Just want to confirm.

MS. MUNTHALI: Everyone agree?

MEMBER SPANGLER: I think Barry-Lewis wants to discuss evidence.

MS. MUNTHALI: So we go to performance gap.

MEMBER SPANGLER: Yes. So, as noted, the performance gap I thought was kind of interesting. Because it looks like from two

different sources that it's not large. But in another context, it seems like it is. It could be up to, you know, 20 percent. So, you know, only 80 percent versus the low 90s.

So there is some performance gap. It might be a little higher than others, but it still exists.

And then when it comes to disparities, there's definitely some disparities in certain populations. I'm not sure, you know, with the racial disparities it says it's statistically significant.

I'm not sure how that corresponds to clinically significant, between 91 and 95. I mean, they're both very high percentages. But in some of the other populations, there's definitely lower percentages in the low 80s, and obviously we want this to be 100 percent. So I would say it's probably, the performance gap, like as you guys stated there, is probably moderate.

MEMBER HARRIS: Exactly what he just stated.

MS. MUNTHALI: Comments, questions.

MEMBER TEUTSCH: Yes, a question though. When you're up around 90 percent, how much better is it going to get? People are in the hospital very short periods of time, they got --

MEMBER SPANGLER: It's a great question.

MEMBER TEUTSCH: I mean, really, aren't we topping out on this one?

MEMBER SPANGLER: I think, it's great question, Steve. I think with other measures if we were at this percentage, we would say, Yes, we're topped out. I don't know if that answers -- but yes, I don't. Because I think, I've been involved in discussions with other measures, and this is where we say we're topped out. But for this, I'm not sure.

MS. MUNTHALI: Do others have thoughts on the performance gap or lack thereof? Arjun?

MEMBER VENKATESH: I don't remember the exact definitions that are used, but CMS's

value-based purchasing program has a couple definitions for topped out.

One is I think based on a coefficient or variation. And then the other one, if I remember right, is you compare the 75th to the 90th percentile, and see how different those are.

I don't see those two here, or maybe I'm not looking in the right place on the form. But maybe that would help give us some general guidance at least.

MS. MUNTHALI: Bob, do you have access to those data?

MR. DICKERSON: The -- yes, the 75th and 90th percentiles are on the evidence to support the measure. 75th percentile was at .9652 percent, and 90th percentile was at .9978.

MEMBER SPANGLER: And there's still a little bit of a gap now. So I guess my question then is, going back to this, and I don't know the exact, it says -- this bullet point here. Yes, you have the second bullet at the top there.

For the current submission, the

developers saw rates for flu vaccinations and noted that in the flu season, that nearly 10 percent of hospitals, one in five indicated they were not vaccinated.

So does that mean at the other 90 percent of hospitals, it was, I'm assuming higher than that. It was in the 90s or --

MR. DICKERSON: Right.

MEMBER SPANGLER: Okay. So there's a small percentage of hospitals that have a larger gap than everybody else. That's what we're saying.

MR. DICKERSON: Correct.

MEMBER SPANGLER: Okay. Do we know anything about those hospitals? Demographics, or is it regional, is it a certain type of, like is it rural hospitals, or something like that? Do we have anything --

MR. DICKERSON: I don't have that information, no.

CHAIR McINERNY: Well, one of the questions that I wonder about is, although the

gap is relatively small, if somehow we decide not to continue to use this measure because we feel that it's topped out, will that then result in hospitals not continuing to pay close attention to this measure?

MEMBER VENKATESH: This is a question related. So when you've got a measure in the -- I'm trying to think of what we can do in the confines of this committee. So if we think a measure is nearing being topped out, we're just a steering committee charged with endorsement of the measure or not.

And so how do you think about things like use and feasibility in here? Is that, We're not going to touch it because that's the MAP's job? Or, because the way I'm thinking about this on the flip side is it's not a no-work measure.

A lot of work happens at hospitals where there's tons of electronic alerts every single time there's a patient hospitalized that says, did you give a flu shot, did you give a flu shot, did you give a flu shot. They're

collecting all this data, there's an infrastructure to that submitting it.

And so it's not -- should that be figuring into my head of, is it worth this performance gap, given the work? Or are we just going to say, hey, look at the numbers, this is the distribution, is there a gap, and that's the extent of what this committee does?

MS. MUNTHALI: So that's a great question, and we're just about to pull up NQF's other, well, its endorsement. But we can -- the committee can recommend this for reserve status.

And that means that when you're seeing a measure like this one where you feel that there's very little opportunity for improvement, there's not much of a performance gap.

But all of the other criterion are very strong. The evidence, which you already said, the testing, feasibility, use and usability. You can recommend that, you know what, we don't want to monitor this consistently. It's still NQF endorsed. Gets a reserved status

labeling.

And this committee, as a standing committee, you have oversight over the portfolio. So you may want to say periodically, you look at the measure, see if there are changes in performance or if behaviors have changed as a result of this measure being out there for, let's say, two or three years after it's been put into reserve status.

It is still NQF endorsed. It doesn't mean we're removing endorsement. But it just means that perhaps we're focusing our other efforts in other areas, where there may be opportunities for improvement or there's more significant performance gaps. I hope that helps.

So what we do, if that's what the committee would like, we would vote on this measure. And the only way that it would end in reserve status, we'd have to go through the whole criteria, is if you voted, let's say low. And so essentially, it's failing or insufficient. And it's failing performance gap.

We would then continue, we'd ask you, would you like to consider this for reserve status? And if you say yes, we will continue then with reliability, validity, feasibility, use and usability, and an overall vote. So I guess perhaps more discussion on the gap.

MEMBER SPANGLER: The only thing I would want to mention, which seems a little bit significant compared to everything else we discussed is, there does seem to be a gap within the, you know, Native American or Alaskan Native population, compared to everybody else.

I mean, everybody else is in the 90s, except for that population. So I don't know if there's anything we can do about that. And I don't know if those hospitals happen to be those hospitals that treat those sorts of patients or anything like that, but I just want to make sure we keep that in mind.

MEMBER TEUTSCH: Yes, I like what you said. And I wonder if we can't begin to A) encourage those that remain at the low end to

continue to monitor it and work towards improvement. And then call for some time to really revisit this.

I mean, maybe you have a regular schedule but, over, you know, three years, five years, whatever the right number is. Because I think his argument's implied. There's a huge opportunity cost from all this stuff.

And there are probably higher priority things than trying to goose these hospitals that are already performing well.

MS. MUNTHALI: So we can -- the committee can recommend that we look at it, you know, after we have some trend data, perhaps after two or three years. Even when we don't have projects to review measures, we have funding to do maintenance work.

That's when we did the updates to the specifications. We worked on the access to care guidance and framework. And so we can look at measures that have been put into reserve status and see if we want to remove that labeling of

reserve status in that period as well. Bob?

MR. DICKERSON: I would like to just mention something in reference to when we talk about differences and statistical significance and clinical significance.

The sample population for this measure, during the time period that reported the data and did the analysis on, was a little over 1.5 million cases. And out of those, about 92,000 were not screened and/or vaccinated.

And if you extrapolate that to the larger population of patients discharged from hospitals during that time, we're talking about a little over a million patients that were not screened.

MS. MUNTHALI: Ron has a -- and Matt, I'm sorry.

MEMBER BIALEK: I would like to follow up on Jason's comment about the Native American population. Are the Indian Health Service facilities and also the tribal self-determination facilities part of this measure and part of the

data collection?

MR. DICKERSON: It would be any acute care hospital that's submitting data. So it very well may include some of those. I don't have the exact breakdown on which hospitals. That is something that for future analysis, we could look and see if that data is available.

MS. MUNTHALI: Matt?

MEMBER STIEFEL: I like the idea of reserve status, and this is how we end up with hundreds of measures, is that measures never get reviewed or retired.

Objectively, looking at opportunity for improvement, this is much smaller than any of the other things that we've seen. So if we're just going by this literal interpretation, it seems like very low opportunity.

Now, if the gaps are for a small subset of the population, Native Americans or whatever, then perhaps there's a measure related to disparities. But not for an overall measure for all hospitals in the country.

MS. MUNTHALI: Jason?

MEMBER SPANGLER: Can you provide some clarification? So, the numbers that we see here, with the disparities, is that from the old measure with pneumonia patients, or is that for -- because I see in a couple places it says overall pneumonia patients.

So I'm trying to clarify which numbers are from pneumonia patients and which numbers are from all hospitalized or all inpatients. Because that would sway my thinking on the performance measurement gap.

Because I would assume that there should be a high -- I'm sorry there should be a low performance gap. There should be, for pneumonia patients, it's more likely I would think that they would get screened and vaccinated versus other patients.

MR. DICKERSON: Right, the 2012 data this referenced, and there is from the pneumonia population, that was prior to the measure becoming a global measure.

The updated data is, so, under the disparities would be the second bullet point. That comes from the 2014-2015 season.

MEMBER SPANGLER: And that's all hospitalized.

MR. DICKERSON: Yes.

MEMBER SPANGLER: Yes, got it, okay. Thank you.

MS. MUNTHALI: Other comments or questions? So again, for reserve status, with NQF endorsement. For that to be an option, you would have to go either low or insufficient, and then we'll have a separate vote that says, would you like to consider this for reserve status. And that's a yes or no.

So, the options are high one, two moderate, three low, four insufficient. And this is for Measure 1659.

(Voting.)

So consensus was not reached on performance gap. So zero highs, seven moderate, seven low, and zero insufficient.

So we will proceed to the next major criterion, which is reliability. Jason, Barry-Lewis.

MEMBER SPANGLER: So again, this is facilities in the specified dates. I thought the exclusions were well detailed, the specifications are good.

One of the issues that they brought up I thought was interesting, it's in the reliability section. It wasn't mentioned in feasibility, I don't know if it applies to both.

And I guess this is something that can be resolved. But they noted that there is a lack of an ICD-10 code for a specific influenza vaccination. So basically, instead of getting this from the code, it has to be extracted from the charts now.

So that would seem to me not to affect reliability so much, but possibly feasibility. So I thought actually the reliability from the specification perspective was good.

The did testing again through kind of

signal-to-noise with the binomial model, and it showed pretty high reliability. I think it was .97, which is very high.

MEMBER HARRIS: I would just concur with that one point. But I was just wondering, does anyone know any information related to the reason for the drop in the code?

MEMBER MOLINE: Do we have confirmation about that? There's 68,000 ICD-10 codes, and some of them are so unbelievably picky and obscure, like, you know, water skiing and hurting your knee while you're a prisoner. But I can't believe that they wouldn't have --. And getting it in in front of me.

Because I've seen the various codes that I've had to try to find when I'm trying to get rid of my meaningful use nasty box, at the top, regarding whether someone's had influenza. Is there confirmation from anyone else? Catherine? I mean you're doing charts all the time.

MEMBER HILL: Yes, I was thinking that

was coded using CPT, not ICD.

MEMBER MOLINE: It's a CPT not an ICD. Because it's a procedure code.

MEMBER HILL: Yes, we pull it up using CPT.

MEMBER SPANGLER: But there wasn't an ICD-9 code for that, so.

MEMBER MOLINE: But if they're pulling up ICD -- if they're pulling up codes, they can pull up a CPT as easily as they can pull up an ICD-10.

MEMBER HARRIS: Right here, it says that the ICD-9 diagnosis V04.81 and V06.6 both will convert approximately to ICD-10 diagnosis code Z23, encounter for immunization. But is that going to be specific for influenza?

MR. DICKERSON: And that's exactly what we found, is while the ICD-9 had a specific code for influenza immunization, for whatever reason, when the decision was made of the transfer of the ICD-10s, there are two general immunization codes for any immunization given in

the hospital setting. They aren't specific to influenza immunization anymore.

MEMBER BAER: So the question would be if the inpatient claim would have the CPT code on it, if the ICD-10 is not specific, does the inpatient hospital claim have the CPT code which would be specific for it?

So I can't answer that question, but maybe that's, you know, something that somebody needs to look into.

MR. DICKERSON: Yes, and my understanding was that the CPT codes are not used for hospital admissions. We can check further into that, definitely double check that.

One thing that we did find when we discovered that there was no specific ICD-10 code, we went back to the previous year to try to identify how many cases were identified for having received the immunization by having an ICD-9 code on their medical record, and it was a very, very small percent. It was, I don't remember the exact number, but it was less than 5

percent.

MS. MUNTHALI: Great, any other questions? Okay, we can proceed with a vote on reliability for Measure 1659. One is high, two is moderate, three is low, and four is insufficient.

(Voting.)

Missing two votes.

Okay, reliability for Measure 1659, seven voted high, two voted moderate, four voted low, and one voted insufficient. So we are just beyond consensus not reached. So this measure passes on reliability. And we'll go on to validity.

MEMBER SPANGLER: So empirical validity testing was done, using the traction, and there's demonstrated a high correlation with both the discharge disposition as well as the immunization status. So I thought overall, it was, you know, pretty good validity.

MS. MUNTHALI: Other comments? Okay. Time for a vote on validity.

So the highest this can receive is moderate, because empirical testing was done at the data element level, not at the measure score level. So one is moderate, two low, and three insufficient.

(Voting.)

So for Measure 1659, validity testing, 11 voted moderate and three voted low, so it passes. So we'll move on to feasibility. And Jason and Barry-Lewis.

MEMBER SPANGLER: Yes, I don't think there are any issues with feasibility. And it seems like maybe we've even resolved the coding thing, so I thought it was pretty high feasibility.

MS. MUNTHALI: Other comments?

MEMBER HARRIS: The only thing I said that this was a chart abstraction, but it's not an automatic from the electronic health record. There's no way for it to be completely from an EHR, because some of the elements could be potentially overlooked by human error versus

automatically generated.

MR. DICKERSON: So I know one of the things that is happening right now for this measure is we are looking at respecifying it as an eCQM.

One of the things that hospitals have done in terms of trying to more consistently apply the immunization screening is they have built electronic screening processes into their EHRs. So we're currently working with hospitals in a testing system to see what information can truly be extracted electronically.

MS. MUNTHALI: Ron?

MEMBER BIALEK: So help me understand, with there not being the ICD code any longer that specifies flu vaccination. Oh.

MEMBER HILL: Yes, I think he's right, there is the Z23 code. The challenge is it's not specific to influenza anymore.

MEMBER BIALEK: Right, so my question is how does this remain feasible without that specificity in the code?

MR. DICKERSON: So in terms of the manual chart abstraction they can identify from other documentation that is not like an ICD-10 code whether or not they received the vaccine in the hospital.

MEMBER BIALEK: So when the testing was done, was it done looking at that process, versus the ICD process?

MR. DICKERSON: The testing for what we have in front of us was based on using the ICD because that was what was available at the time the testing was done. Yes, the ICD-9.

MS. MUNTHALI: So perhaps I can help here. So NQF, our policy is that we wanted by October of last year, when ICD-10s were implemented, for developers to at least show a crosswalk between ICD-9 and ICD-10, recognizing that it would take quite a few years for uptake and for developers to have access to test beds to do that.

So there is some lag time in between, as long as in your measure submission, which I

think you've provided a crosswalk between ICD-9 and ICD-10, and perhaps where those, there is a line between those codes that would meet our requirements.

MEMBER BIALEK: I think that the norm though, is that codes are added. And I could see that argument. In this instance, a code was removed, which potentially adds a substantial burden to the facility.

CHAIR McINERNY: But if the patient gets an influenza immunization, then that would be a CPT-4 code, right? And that would be feasible to collect that.

MEMBER HILL: Where it's embedded --

MS. MUNTHALI: I'm sorry, can't find the mic.

MEMBER HILL: Three, oh, yes. So the question is where is that CPT code embedded. And if they've been querying a different level of code in the hospital because of the DRG influences, you roll up the ICD-10s to get the DRG. Then that would be problematic.

MEMBER HARRIS: But the CPT codes for this are two, four, six -- there are six different administration codes for vaccines, and it's not specific for influenza. So that's not in 04604614714727374.

MEMBER TEUTSCH: There must be a code somewhere for the specific drug, or in this case, biologic, that's being administered, as opposed to -- because the CPT will capture the administration, that's the procedure, right? But not necessarily the drug.

And there are, I don't recall how all those drugs are coded, but there should be a code for --

MEMBER HILL: There are HCPCS codes for the drugs.

MEMBER TEUTSCH: Drugs, yes. So it should be somewhere.

MEMBER HILL: Yes, CPT codes for the administration and then the associated diagnosis. And you go to figure out whether you're going to get paid, you go to the national, look at the

national or local determination codes to see how many of those you have to have to get --

MEMBER SPANGLER: The other point, Arjun made this point earlier about how many people are actually going to use the CPT code in the hospital. Because if it's part of the DRG, you know, I mean, it just. Why are they going to do it, yes.

MR. DICKERSON: Our understanding is the CPT code is not used in the hospital.

MEMBER VENKATESH: I guess I wouldn't stress too much about the coding, and the reason is this: the purpose of the code for the flu vaccine is to identify cases for the numerator, right, not the denominator. The numerator for this measure is already sky-high.

So, even in the current world of however this data's captured, and it's actually captured by chart abstraction. Mostly people are looking through the chart to see was this patient eligible for a flu shot and was it given or not. Performance is already well north of 95 percent.

And so whether or not you capture a few that are coded in procedures here and there actually doesn't matter. I'd worry a lot about this if performance was low and we said, Oh, well, this was only happening 30 percent of the time, and it's probably because it's not being captured in claims. But they are capturing. They're finding the numerator right now.

MEMBER SPANGLER: It's obviously feasible because of the percentages.

MEMBER TEUTSCH: I mean, it's feasible, I mean, we know you can get it from charts. But the problem is it's labor-intensive to get it from charts. And you know, if you could get it from codes, then you say, Well, that's pretty simple to do. But to me, this is an enormous burden for relatively modest gain.

And so it depends what we mean by feasibility. Is it possible? Sure it's possible, but I think this is a practicality question, isn't it? As much as anything. And somehow I think our job is to assess the

tradeoffs.

MEMBER HARRIS: So that was just my point when I asked about the EHR component, because I think it's just a lot of work that would go into capturing this information. And specifically, with the translation of the coding, it would just, it's going to make it even more onerous or laborious.

MS. MUNTHALI: Amy.

MEMBER MINNICH: I'll just come out of my shell. From a clinical and formatic standpoint, there is technology that could help with some of that unstructured data, such as natural language processing. And so there are other efforts that you can put in place to try to get that data.

MEMBER SPANGLER: I think what Steve's asking is the definition there of could it be captured without undue burden. Because that's what the definition is there, feasibility. And I think that's what you're asking, is it an undue burden or not.

MS. MUNTHALI: Matt.

MEMBER SALIVE: Is this issue only applied to this one of the nine flu measures?

MS. MUNTHALI: This is a --.

MEMBER SALIVE: Can we start over?

MS. MUNTHALI: This is a great question.

MS. MUNTHALI: Yes, it is, it is. Now this is the first time you guys brought it up as a significant concern. But perhaps you have some data you can share. Bob, I don't know what you've learned from implementation of this. Well, you've heard from end users on the implementation of this measure to share.

MR. DICKERSON: In terms of abstraction burden -- or, okay, just wanted to make sure I was clear. Actually, most of the chart-abstracted versions of measures, this one probably has the least amount of burden associated with data collection, because of the way that hospitals are collecting the data in screenings forms when patients arrive to the

facility.

They'll actually have screenings where a nurse has this built in a part of their admission process, and they check, has the patient received it before, are they eligible for it, refuse it, have a contraindication. So a lot of that information is captured up front in a screening form.

And then for the folks that are doing the abstraction, they can go to that information. And if one of those options is not checked, it really narrows down what they need to look for.

MS. MUNTHALI: Okay, I think we'll move to a vote on feasibility for Measure 1659. One is high, two is moderate, three is low, and four is insufficient.

(Voting.)

Okay, on feasibility, two voted high, eight moderate and four low. And so this measure passes on feasibility, and we'll move to usability and use.

MEMBER SPANGLER: So pretty high

usability and use. It's already being used in several publicly reported programs already.

MS. MUNTHALI: Barry's giving the thumbs up. Any other comments, questions? Okay, we'll proceed with a vote. For 1659, one high, two moderate, three low, and four insufficient.

(Voting.)

Steve?

MEMBER TEUTSCH: Again, it gets back to the definition here. Is it going to be used for accountability and performance improvement? If we think that improvement can't go anywhere, then isn't that low? I mean, usability, yes, it's practical and people are using it.

But at what point do you say that it's not much use for performance improvement?

What do we do to repeal the vote? Take it to the legislature?

MEMBER VENKATESH: I think what people are struggling with, I'm struggling with this as well, is like when you look at usability and use, is it kind of a retrospective assessment? So

highly successful are going to have had a lot of use and improvement, right. Like part of this is, was there improvement since you used the measure.

And there has been since this got publicly reported. If you look at that data, the mean went up like 15 percent, 20 percent. Or is this a prospective guess from this committee? Committee on like what's going to happen in the future. Because then agree, I highly doubt that the performance is going to improve from where we're at right now. I don't know how you -- I'm sure this has come up.

There have been plenty of measures that have been near topped out that committees have struggled with this issue before. And so I'm just thinking like we should at least try to consistent within this group. Because we'll see more measures that are topped out too.

MS. MUNTHALI: Yes, and this committee, I can tell you, for many of the other topical areas, like surgery and safety and

cardiovascular, well, we've had measures in our portfolio for quite some time. A lot of them, of the standing committees, have been opting for reserve status. Because they've had some sort of trend data over time.

They can see that you know, there have been improvements, but there's really not that much room for additional improvement. So you are starting that discussion as part of your committee. And I can't remember any of our measures in reserve status.

So I can see maybe three years from now, quite a few of them may be in reserve status. So this is why we offered it as a recommendation for you guys, if that's the option you wanted to take. Matt and Barry-Lewis.

MEMBER STIEFEL: Can we vote for a reserve status even if it passed all of the other criteria?

MS. MUNTHALI: It has to pass all the other criteria for you to vote on reserve status.

MEMBER STIEFEL: Oh, okay.

MS. MUNTHALI: Because we're saying it's still a sound measure, it's just one that, you know, we may not be monitoring as frequently as a measure that is endorsed without reserve status.

MEMBER HARRIS: Yes, so that was actually what I had turned the chin up was for, how can we place it, you know? On the agenda as a reserve, you know, for the next cycle now, so that we don't get to next time and revisit this whole.

MS. MUNTHALI: So what would happen is on performance gap, where we have consensus not reached, it was seven moderate, seven low, yes, seven low. It's consensus not reached. You may want to think about it, receive the comments from NQF members and the public during the post-comment call.

That may sway you to say, You know what, we really think this is a sound measure but there is very little room for improvement. We want to monitor it, but not as frequently as we

do others. So there's still an option. But it has to pass all of the other criterion.

Okay, I think Matt and then Bob.

MEMBER STIEFEL: How do we do that?

MS. MUNTHALI: How do you vote on it?

MEMBER STIEFEL: I mean, does it take a motion to put it in reserves?

MS. MUNTHALI: Yes, but we have to finish. We've done usability and use, so we can go back to performance gap if you'd like, now.

Or if this conversation, since we talked about performance gap, has swayed you to insufficient to low, then we would just decide as a committee to say, Yes, we would like to recommend this for reserve status or no. If you vote no, the measure is not recommended. So.

MEMBER SPANGLER: So how many have to vote low or insufficient to get to that point?

MS. MUNTHALI: So you have to have 61 percent.

MEMBER SPANGLER: Oh, so it's the same as for the passing things you have to have that.

MS. MUNTHALI: Exactly.

MEMBER SPANGLER: So the negative it's a, wow. Okay.

MS. MUNTHALI: So that's something you should consider.

MEMBER SPANGLER: But do we need a motion to go back and revisit that, or are we going to do that?

MS. MUNTHALI: Do you want to make a motion? Somebody from the committee should make a motion.

MEMBER SPANGLER: We need to reconsider the previous vote, is what your Robert's rules would say.

MS. MUNTHALI: Okay.

MEMBER HARRIS: We would move to reconsider --

CHAIR McINERNY: Performance gap.

MEMBER HARRIS: Performance gap motion.

MS. MUNTHALI: Okay, so yes, we'll finish use and usability. But wanted to also

remind you that you did ask for some specific performance gap information on Native American populations and other sub-populations.

Do you think that's significant enough that would say, Well, you know we probably should endorse this measure without condition of reserve status. So I just, if that's the case, if you think that's going to be significant enough, then let's wait until the post-comment call and see what Bob and his team can produce for us in the way of those data.

Then we proceed. But if you think that's not significant enough to sway you to vote for reserve status, then we can continue with that motion.

CHAIR McINERNY: So first of all, feasibility looks like it passes. What's the percentage?

MS. MUNTHALI: Usability.

CHAIR McINERNY: Oh, usability looks like.

MS. MUNTHALI: It's, we're good on

usability and use.

CHAIR McINERNY: Yes.

MS. MUNTHALI: But then Steve was concerned about where we factor in opportunities for improvement.

CHAIR McINERNY: Yes, so the question before the committee is do we want to re-vote now on the performance gap? Yes.

MS. MUNTHALI: And before we do that, Bob, you had something to say. Sorry.

MR. DICKERSON: Yes. So one of the things when we're looking at performance of the measure, and yes, in terms of percent of patients that are being screened to receive the vaccination, it is in the mid-90s.

Also of note is when we look at this from a hospital perspective, the percent of hospitals that are scoring less than 90 percent of their patients being vaccinated is in the range of 20-25 percent. And the reason it's 20-25 percent is that analysis was done based on the reporting quarter.

So for fourth quarter of 2014, it was 25 percent, and for first quarter of 2015 it was, let's see, actually about 16 percent rather.

And I think that's one of those areas where, regardless of what the decision is of this committee, I think it merits some additional analysis to try to identify why do we see that large a percent of hospitals scoring below 90 percent but we're seeing 94 percent roughly of all patients immunized and screened.

And then, as it was pointed out also, the disparities, the Native American group and others.

CHAIR McINERNY: All right, so are we ready to re-vote on performance gap?

MS. MUNTHALI: And just to clarify, 61 percent of you would have to vote low, because you're saying there's no opportunity here, very little opportunity for improvement. Not insufficient but low.

(Voting.)

So it is 64 percent. So we can take

a vote on whether or not you would -- so let me read this out for the record. Sheila can bring up the votes.

So zero voted high, five voted moderate and nine voted low for performance gap. And so the committee would like to consider whether or not this measure can be placed in endorsement with reserve status.

And so Sheila, if you can bring up, I think there's a slide for that, if there's not. I can't read the fine print. Maybe you should read it out. Sorry.

MS. CRAWFORD: Endorsement maintenance, potential for reserve status. If a measure is under endorsement maintenance review and did not meet importance to measure and report only due to lack of performance gap 1d, does it meet criteria to consider for potential reserve status?

High performance is likely due to actual improvement versus issue with measure construction. Strong direct evidence proximal to

desired outcome, high ratings for reliability and validity, possibly moderate, demonstrated use, demonstrated improvement. One yes, two no.

MS. MUNTHALI: Thank you.

MEMBER SPANGLER: Do we have to, this supplements voting for endorsement by itself. We don't have to vote for endorsement and then do reserve.

MS. MUNTHALI: Yes. You just do this.

MEMBER SPANGLER: Just do this, got it.

MS. MUNTHALI: So if you say no, majority 61 percent and over say no, this measure would not be recommended for endorsement. Because remember, it received a low vote on performance gap, which is a must-pass criterion. So, again, just wanted to put that there. So if you wanted for reserve status, you should say yes. If not, say no.

(Voting.)

So one more vote. It's unanimous, 14 voted yes for endorsement with reserve status.

So it retains its endorsement, it's NQF endorsed. And as we mentioned before, we'll be looking at it periodically.

CHAIR McINERNY: All right. Congratulations, we did something relatively new. And more importantly, we're done with flu.

(Applause.)

MS. MUNTHALI: Okay, we're going to call up our colleagues from PCPI Foundation for Hep C measures, and we're going to go through this we hope rather quickly.

MS. OGUNGBEMI: Hello, may I have your attention please. I would like to ask if you all would still like to go to dinner tonight. It would be at P.J. Clarke's, which is like a burger and seafood restaurant about a block away.

Our reservation is for 6:15, so if I could get a head count. We'll be done. We have to be done. Yes. I could try and change the reservation, but we can see.

CHAIR McINERNY: By the way, P.J. Clarke's is a great restaurant. Thank you for

picking that.

MS. OGUNGBEMI: You're welcome. Now the reservation is at Georgia Brown's because P.J. Clarke's does not have any -- I'm not sure what happened. They don't have any tables.

MS. MUNTHALI: Georgia Brown's is good.

CHAIR McINERNY: Yes.

MS. MUNTHALI: So we'll make a reservation for about ten, maybe. Staff is going, and we'll make it for ten. Whomever shows, shows.

CHAIR McINERNY: 6:15, Georgia Brown's. Where is it?

MS. OGUNGBEMI: We'll give you the address.

CHAIR McINERNY: Okay, thank you.

MS. MUNTHALI: Okay. So the next set of measures, the three measures that assess an aspect of Hepatitis C screening, Measure 3059: One-Time Screening for Hepatitis C Virus for Patients at Risk. And the other is an Annual

Hepatitis C Virus Screening for Patients who are Active Injection Drug Users. And the third one is Appropriate Screening Follow-up for Patients Identified with Hepatitis C Virus Infection. So, we asked PCPI Foundation to go through these rather quickly because they flew in from Chicago. They can't come in tomorrow. I'm going to try to get through this very quickly but they're going to summarize the three measures together and we're trying to find other efficiencies as well.

Thanks.

MS. BOSTROM: Thank you. My name is Beth Bostrom of the PCPI Foundation and I will kick it off to our expert Dr. John Ward.

DR. WARD: Good afternoon, everyone. I'm Dr. John Ward and I'm the Director of the Division of Viral Hepatitis. At CDC we developed these measures in collaboration with PCPI. It's very appropriate that we talk about all three of them together because they're striking to the heart of a key prevention priority at CDC and the USPSTF and that is to increase testing and

knowledge of HCV infection among persons at risk and link them effectively to care and treatment at a time where we now have safe and highly effective cures for Hepatitis C.

A quick survey that was in your materials that you read prior to this meeting, the burden of Hepatitis C is large, about three and a half million people are living with Hepatitis C. The knowledge of infection is low, only about half the people based on NHANES report knowledge of their infection before they were tested based on that survey. Mortality is increasing. It's now the largest cause of infectious disease related mortality among conditions considered to be nationally notifiable to CDC. And, in fact, you put all the other 60 notifiable conditions together and the number of deaths from Hepatitis C supersedes all of those combined. And those stats are among primarily among this birth cohort that has the high prevalence with an average age of death at 59. And there's large disparities within that cohort,

particularly, high prevalence and mortality among black Americans and among American Indians.

The other area that we want to address is an epidemic of HCV transmission. There's been about 150 percent increase in incidents between 2010 and 2014. This is among persons who inject drugs and we have a none disparity here where it's almost predominantly white in suburban and rural areas of the country equally male and female populations that have not experienced this type of transmission. And this is very much related to the opioid epidemic and then the injection of those opioids leading to Hepatitis C transmission. And we're now seeing even some secondary transmission from mother to child.

The benefits of testing. There are some benefits of knowledge of your status and also in changes in care and behavior such as changing your alcohol use, which can lower your risk of liver disease but the linkage into treatment is very, very important we now have drugs with one to several pills a day for eight

to twelve weeks, 90 percent cure in clinical trials and the data show very little drop off in that success rate in routine clinical practice.

But people who are cured of their Hepatitis C they have large reductions in liver cancer rates, all-cause mortality and progression of their cirrhosis so there's some clear benefits.

Since we prepared this packet for your consideration we have continued to publish data to show how the implementation of these testing policies and linkage to care can increase the number of people being successfully linked to care among low income African Americans in Philadelphia, Hispanics being cared for and safety net hospital in San Antonio, Texas, and among the American Indian populations in the Indian Health Service and in the Cherokee Nation.

We have a large gap in testing and linkage to care that we hope that we can work together with NQF to address and measure and improve performance.

Thank you.

MS. MUNTHALI: Thank you, Arjun. Amelia, I don't know if you've joined us. I think it's just Arjun.

MEMBER VENKATESH: I guess one question because I reviewed a bunch of the eMeasures. This is an eMeasure for full endorsement or for the trial standard use?

MS. MUNTHALI: Trial use.

MEMBER VENKATESH: Trial use. Okay.

So, this measure is a great description obviously a very important topic. It is essentially a high-risk Cap C screening measure and I think in my review of this as a trial use measure the idea here is that we evaluate whether or not there's enough here to suggest the problem is important, that there is some practice that we should measure this and that they should kind of fully develop the eMeasure and then do some of the subsequent validity and reliability testing.

The screening types of analysis that

were done about this in terms of measuring properties to me all looks good in terms of being able to capture a lot of this data electronically. I think that all makes sense. The only concern I had about this measure that I put in my original review and I don't know how and where to put it in this discussion is not actually about the electronic specification or about the measure. It's about the evidence and the performance gap.

On the evidence side there is primarily the reason to rate this probably moderate is that there is a lot of clinical practice guidelines and consensus things including from the CDC that suggest that this is important and that high risk people should be screened. And I can live with that and you can rate that as high as I think moderate based on what they've done.

The issue becomes on the performance gap there's two studies cited about inadequate screening. And I did not do the extra, you know,

literature review to find other papers. But the two that are cited come from NHANES data between 2001 and 2008 is one study that suggests that people don't know their Hep C status. That's an important gap and problem that's different than this measure. This measure is not about patient knowledge. It's about patient screening so the second study they cite is also from NHANES data 2006 to 2008. In that data of the 12 percent of people that did not know they had Hepatitis C, I think five and a half percent tested positive on the blood sample. So that to me would be good data except I'm a little afraid that it's almost a decade old looking at whether there's inadequate screening. And so I am left with I think as I looked at the evidence and performance gap on this measure something I think is really important, something where there's strong consensus about the importance of screening but I don't necessarily know that there's a gap that suggests that there is -- I don't know that there is evidence that shows that there's a screening

gap right now. And I just don't know where that fits into a trial use measure. Because it's not about the measure and the eSpecifications. It's just about the evidence of a performance gap right now.

MS. BOSTROM: And I think if we could speak to Dr. Venkatesh's point just a little bit. I believe Dr. Ward touched on this a bit in his introduction but we do have some newer data from Indian Health Services that includes 1.9 million members, that includes 566 Federally recognized tribes, through a wide network of facilities that did implement a performance measure that really looked at cohort screening, one time cohort screening for those at risk and what they did find was that from 2012 to 2015 the baseline rate increased from 7.9 percent to 32.5 percent. So, I think that the newer data definitely supports that. And the study also showed further variation. More women received screening than men and that there were also some geographic variation. Regions varied from 31.2 percent to

41.2 percent and there was even wider variation amongst facilities measuring from 1.9 percent to 41.2 percent. So, we thought that that could further highlight, you know, some more recent efforts about screening for the at risk population.

MS. MUNTHALI: Jacki.

MEMBER MOLINE: I just have a quick question. If someone had been screened five years ago it looks like it's looking for current screening as opposed to past screening. So, can you just clarify if the numerator or denominator because if someone's in the baby boomer cohort and he is in the group that's supposed to be screened just solely based on birth date but they were screened five, ten years ago are they included, excluded?

DR. WARD: It should be excluded.

MS. APURA: So, actually our numerator in the specification says that all HCV laboratory test, HCV antibody test, HCV RNA test, HCV Riba Test all those happen before the measurement

period of those included in the numerator and also tests that are included during the measurement period too.

MS. MUNTHALI: Steve.

MEMBER TEUTSCH: Sean, you and your colleagues have done a great job in bringing this to everybody's attention. But sort of the elephant in the room, I'm not sure it's eligible for us to think about and that's the cost.

As these are costly drugs, even though I know there's cost-effectiveness studies out there that show that it's been cost-effective but it's really not affordable for many, many, many and particularly Medicaid, public health clinics. I know in LA County they said if they followed this it would cost more than their total annual budget. And there are other strategies as you know that we could use so that you don't have to take everybody all at once. People can get reinfected. Even if they are treated if they keep up some of the high-risk behaviors or you eradicate this disease. It's not necessarily

just a one-time cost.

So, I don't know and I guess it's an NQF question. Are we supposed to think about things like that as a trial measure to see if you can get it? Does it meet specifications? I think we've heard that that's all good. But I truly worry about really the affordability vis-a-vis the system more broadly. The Government could technically march in and take over those patents because they have march-in authorities and they've chosen never -- not just to this but they've never used those abilities. So I just wonder though, you know, given that it's a real public health problem you've laid out and there's some really good solutions from a medical perspective but we have a real mismatch with the economics.

MS. MUNTHALI: So -- no, you definitely should respond. Yes, cost is something that we do consider as part of our feasibility assessment of every measure that comes through regardless if it's a measure that's seeking endorsement from

NQF or a trial use. Trial use is not endorsement. We are saying, you know, these measures should go out there. The developers should go out and test as best that they can. They have three years to come back to us. Some measures may not make it and it could be because of the barriers with testing or it could be some of the resource barriers like cost. We don't know that but I just wanted to put that disclaimer on there again that this measure and these other that will proceed in this group they're coming forward for trial use not for NQF endorsement. So, I don't know if that helps.

MEMBER TEUTSCH: Well, it helps me a little but I believe that when they're done you can show that this is feasible. I don't have really much doubt that this can be done. But then you're sort of you're asking the same question. The question is when do you ask the question as to, you know, you're sort of on the slippery slope, if you will.

MS. MUNTHALI: You brought up

feasibility issues. And that we talk about anytime with any measure.

DR. NISHIMI: It's not just the technical feasibility of it. It's whether the cost of doing the measure outweighs the benefits too. So, it's not just, you know, we can collect this data and give you a score.

MEMBER TEUTSCH: My guess is, it's feasible and cheap to do the measure. It's not the measure that's the problem it's the cost of the drug.

DR. WARD: Let me chime in at this point.

Within the testing and the linkage to care that everybody should know their status in care so that their stage of disease can be evaluated. When you look at the birth cohort and we've done a study and Stanford has done a study about one out of every four baby boomers here have Hepatitis C infection already had severe fibrosis or cirrhosis. So, even in the most restrictive Medicaid reimbursement criteria they

would qualify for treatment because it is so highly cost effective and probably cost saving to treat people at that severity of disease. But that's one thing.

Two, as far most baby boomers, I'd say about 90 percent of the baby boomers are transmission dead enders. You know, they're not engaged in risk behaviors to transmit to others. They were infected decades ago and now they're getting ill.

On the cost side, the original drugs that came out in '14 were about $86,00 to $94,000 per puritive course. Through competition and negotiation in just two years that's fallen to about $46,000. So, our original cost effectiveness modeling on a societal level showed that that was cost effective to treat people who were infected. We just updated that analysis with using a benchmark of about $41,000 and showing that it's cost neutral now because of the benefit of the reduced cost. So, I think we have to look at the true cost as a moving target. I

mean, I think this is an evolutionary process and it's heading in a positive direction in just two years. But we need to be mindful of cost, but again I think what we're measuring is the testing and getting people into care so that they can be staged and then see what is their priority for treatment. As you were saying, you may not treat everybody immediately. You may not. But they all need to be in care and staged to make that decision. That's the way I look at the three measures collectively.

On the injection drug use, quickly on the injection drug use side they do have risk for ongoing transmission and I think there's a different cost effectiveness that we have to put into play there and we're still trying to figure that out.

And then lastly it's very interesting you should bring that up, Steve, because the other reason I'm up here is because HHS has convened a two-day meeting to look at how they can improve the affordability of HCV treatment.

So, there's different options for that.

Thank you.

MS. MUNTHALI: Other questions? Concerns?

Okay. So, we'll vote on Performance Gap, one, high, two moderate, three low and four insufficient.

(Voting.)

MS. MUNTHALI: Sorry, I'm reading what's on the screen.

So, this is a new measure. It is eligible for trial use not NQF endorsement but we still need to assess the evidence base. This is Measure 3059.

Okay. Evidence, one high, two moderate, three low, four insufficient.

(Voting.)

MS. MUNTHALI: So, we have 13. Ron is gone for the day. So, we have four high, eight moderate and one low. So, this measure passes on evidence so we'll move on to Performance Gap now.

DR. NISHIMI: Arjun, you were the one

who raised the issue about Gap. Is there anything you wanted to add? You've heard the PCPI Foundation response with the updated information.

MEMBER VENKATESH: I guess I would still interpret it probably as -- I guess how do you for these trial standard measures how do you interpret a gap that's not specific to like the measure focus?

DR. NISHIMI: Well, wasn't your Indian Health Service specific to the focus of the measure?

MS. BOSTROM: It was specific to the birth cohort.

DR. NISHIMI: Well, I mean it's really up to the committee whether you feel that by inference you can feel comfortable voting moderate that there is a gap. It is by inference.

(Voting.)

MS. MUNTHALI: We're missing one vote.

(Voting.)

MS. MUNTHALI: Okay. Three high, seven moderate and three low for performance gap for Measure 3059. So, it passes and we move on.

DR. NISHIMI: The next section is on the Scientific Acceptability. Recall that there's no testing data here because this is for trial use so the Committee's decision really is about whether you feel the specifications are appropriate given the evidence.

MS. MUNTHALI: Arjun, did you have any other concerns with the specs?

MEMBER VENKATESH: No, and on the sheet it had a big red don't evaluate.

DR. NISHIMI: Well, the evaluation really is around whether there is evidence in the specs you feel match the evidence.

MS. MUNTHALI: Right so you're not voting on reliability validity.

DR. NISHIMI: So, we ready to vote on scientific acceptability? We're voting now on the scientific acceptability because we're not voting on reliability per se. We're not voting

on validity per se.

(Voting.)

DR. NISHIMI: Oh, okay. Then this is for Measure 3059. It's not for reliability. It's for the scientific acceptability and I can't see the number.

MS. MUNTHALI: So, four high, eight moderate and one low so it passes. So, let's move on to feasibility. And there should be some discussion here. Arjun.

MEMBER VENKATESH: I mean what they submitted in the original application suggested that the preliminary feasibility looked pretty good. I don't have any reason to think that you couldn't get a lot of these things from the HR. I guess they'll show us when we see it the next time around whether or not -- the main concern would be whether or not you're missing people who have a history of a one-time test. So, if you move around, if you capture one of the HRs the data is sitting somewhere else but that seems like it's outside the scope of what we're trying

to figure out today. And so from an overview look at feasibility it seems like these data elements are in EHR and I'd give it, I guess moderate? High?

MS. MUNTHALI: Steve.

MEMBER TEUTSCH: Can I ask John, and Colene, so feasibility is really about not just the performance metric per se but sort of, you know, are people going to use this and to what extent to actually drive change in the organizations? Can you talk a little bit about how different types of organizations whether they're plans, HMOs, you know, public health clinics, public health hospitals are actually prepared to use, you know, to do this? I know there is a philosophical yes, we'd love to do it. But on a practical side what are they actually doing in terms of the actual, you know, implementing and then the screening and then, you know, the management of those people?

DR. WARD: So, we're tracking testing using millions of records from two large

commercial laboratories. We've seen 60 percent increase in testing since these recommendations were put forth in 2012. And that's with a very limited implementation budget from CDC.

And you saw an example from the Indian Health Service of a system really finding ways to implement this through provider education and clinical decision tools so that you get pop-ups when someone enters into the clinic, in the birth cohort. So, and there's been endorsements by America's Health Insurance Plans. Kaiser has put in a clinical decision tool. So, I think there's a lot of -- there's a variety of effort not as large as frankly I would like but there's some good examples of a major health system doing something and, of course, the VA has been about this for almost 20 years now and have 80 percent. They believe they've found 80 percent now of Hepatitis C veterans who use their veteran system and --

MEMBER TEUTSCH: What about Medicaid plans?

DR. WARD: Medicaid?

MEMBER TEUTSCH: Medicaid.

DR. WARD: Well, Medicaid we are just now getting into looking at that. We have seen increases even among Medicaid populations in testing. You know, one of the big barriers has been getting access to treatment because of the budgetary issues for Medicaid programs. And that's been a disincentive to test. So, that's a problem.

Our big effort going forward, we just put out money to State Health Departments to help them begin to work with FQHCs which we think will help us address some socioeconomic and racial and I think disparities. Many of those will be on Medicaid to improve testing. So, we'll find out more as this moves along. But we have examples of systems implementing testing and testing increasing nationally if that answers your question.

MEMBER TEUTSCH: And then they're following up and treating?

DR. WARD: They are. We have in addition to the Indian Health Services there's a number of other examples of people being linked to care from emergency departments, hospitals, homeless clinics, etcetera. So, it does improve linkage to care even in these budgetary issues.

MS. MUNTHALI: Any other comments? Okay. We're ready to vote on feasibility for 3059, one high, two moderate, three low and four insufficient.

(Voting.)

MS. MUNTHALI: One high vote, ten moderate votes, two low and so this measure passes on feasibility and we'll move on to usability and use. Any new concerns to bring up, Steve, your concern about cost will apply it to usability and use as well. Anything else to add?

MEMBER TEUTSCH: No, I just think from this conversation clearly there's a push to make it work and there's a huge barrier in terms of affordability which makes it really, really

tough. And I know people are working on it but time will tell. And I know just from my local experience that --

DR. WARD: As a brief addendum I mean the USPSTF Category B does help address some of the cost issues of testing because that should be not --

MS. MUNTHALI: Steve, your microphone, sorry.

MEMBER TEUTSCH: Yes, but health claims are supposed to provide that testing service without copay.

DR. WARD: Correct, but they don't in their decision --

MEMBER TEUTSCH: No, I meant in terms of the patient burden.

DR. WARD: Correct.

MEMBER TEUTSCH: Not always but often times that has been addressed in particularly primary care settings.

MS. MUNTHALI: It looks like there are no other concerns with regard to usability and

use. So, we'll start to vote. One is high, two is moderate, three is low and four is insufficient.

(Voting.)

MS. MUNTHALI: So, for Measure 3059, usability and use, one high, eight moderate, three low and one insufficient. So this measure passes usability and use.

And the recommendation for approval in the Trial Use Program. One is yes and two is no.

(Voting.)

MS. MUNTHALI: Eleven voted yes and two voted no for inclusion in the Trial Use Program so this measure is approved.

So, we'll move on to the next measure also stewarded by PCPI and this is 3060. We are not going to do the developer overview. We did that already so just turn it over to Jason, is it you?

MEMBER SPANGLER: I mean I'm in this but I'm not initiating the --

MS. MUNTHALI: Okay. So, it's Arjun

and Betty Lewis, right?

MEMBER SPANGLER: I will join in the discussion.

MS. MUNTHALI: Okay. We have to vote on evidence. It's a new measure. Again, this is an eMeasure that's eligible for the Trial Use Program.

MEMBER VENKATESH: So, I think the denominator is obviously slightly different in this measure but otherwise it is essentially constructed the same. They did the same degree of pre-testing around the data elements. I don't think I had anything new or different on this measure from the previous one. I think that, you know, largely this is going to be based on guidelines to get the evidence rating, not on actual evidence and the performance gap I'm assuming that the additional information that they brought is also true for this measure. I don't know if they looked at the subset of folks with active injection drug use in the Indian Health Service. My guess would be is that

there's gaps in screening in non-IVDA patients. It's also true there as well but I guess that would be the only question but otherwise to me, I think that this is essentially the same general construct as the other measures, the only difference being the denominator.

MS. MUNTHALI: Any other questions?

MEMBER HARRIS: I just had a question about the survey of the low income position. Did we look at any other group or was there a particular reason why that was the place for the survey -- did you like with, you know, the baby boomer population did you actually think about other groups than just that particular low income group?

DR. WARD: So, is your question related to injection drug users or related to birth cohort?

MEMBER HARRIS: Yes. I mean, I don't think that it necessarily means that they have to be a low income group to be an IV drug user.

DR. WARD: No, that's exactly right and

we -- we have directed our prevention research to the settings that we're serving populations that were experiencing health disparities and so that's why it's a little bit directed in that way. But if you look at larger health systems like, you know, Kaiser which is published in the Mid-Atlantic region, for example, you'll see similar rates to what has been shown in NHANES.

Relative to the injection drug use population, I mean, we would like to -- I think the performance gaps outside of drug treatment is a little bit less well described. But, you know, given the behavior we're concerned, and the changes in the geography of poor transmission is happening, we're concerned that there are gaps there that we have yet to fully described.

MS. MUNTHALI: Jason.

MEMBER SPANGLER: That kind of answered my question because it doesn't -- there's not data on disparities in the population we're looking at here though it sounds like specifically. Do we know of disparities of those

who are screened who are IV drug users it sounds like not yet.

DR. WARD: We do get data from the National HIV Behavioral Surveyor, NHBS, which goes out and interviews persons at risk for HIV including persons who inject drugs. And there are racial and ethnic differences and actually with Black Americans reporting higher testing rates for Hepatitis C than White Americans and it's believed it's because prevention services have been in Black communities and urban areas longer than in White communities where they may even be nonexistent, particularly now in the suburban and rural areas. So, there are some data to suggest that there are some performance gaps.

MS. MUNTHALI: Okay, it looks like -- oh, Matt.

MEMBER STIEFEL: Well, the numerator is different too, obviously. It goes from one time to annual, which does affect our thinking about the evidence and the opportunity and especially

the usability or feasibility cost issues would seem to be dramatically amplified. So, I wouldn't want just with a broad brush to say, well, it's just a different denominator.

MS. MUNTHALI: Arjun.

MEMBER VENKATESH: Yes, that's a good point. I should have brought that one up as well. I briefly looked at the guideline and I know that the cited evidence from it obviously suggests, I think the words are regular screening or something like that. Is the one year requirement lined up with the current guideline where it says annual testing and annual screening, because I think that would be reassuring?

MS. BOSTROM: Yes, the guidelines which we provided, the AASLD guideline does recommend annual screening for persons who inject drugs. And then in the USPSTF recommendations it recommends periodic screening for at risk populations and it says the population which is at greatest risk is persons who inject drugs.

So, during the development of the measure the work group felt that periodic would be a feasible way to align with annual. So, it is explicit in the evidence attachment.

MS. MUNTHALI: Okay, great. I think we can vote. We still have a quorum. I think it's 11. Two people stepped out.

So, for performance gaps. We haven't voted on evidence, sorry, again. One high, two moderate, three low and four insufficient.

(Voting.)

MS. MUNTHALI: Okay. So, eight voted moderate, two voted low and one voted insufficient. This measure passes on evidence. And now we'll move to performance gaps. Any additional comments? We started talking about gap.

MEMBER VENKATESH: I guess I just ask the question, you guys had a lot of gap data for a variety of populations on the last measures there. Is there similar stuff you know about this population or was it from the HIV survey the

primary data point?

DR. WARD: We have done demonstration projects in about 10 injection drug use settings and there was a wide variety in their ability to successfully implement testing and wide variations in linkage to care. I can't describe as of right today what's all the reasons for that. But we do have some clinic data to show that there are differences across clinical settings.

Some of that data was just recently published in Public Health Reports in July and I'm happy to make those data available with PCPI.

MS. MUNTHALI: Anything else? Okay, let's vote on gap. One high, two moderate, three low, four insufficient. And we're looking for 12 votes. Measure 3060.

(Voting.)

MS. MUNTHALI: So, two voted high, eight voted moderate and two voted insufficient so we move on to reliability, but testing, essentially, or ---

DR. NISHIMI: This is a specification. So, this is the same issue. It's trial use. So, there's no testing data and the question to the committee is whether you feel the specifications represent the evidence well enough to go forward with the trial use.

(Voting.)

MS. MUNTHALI: Steve.

MEMBER TEUTSCH: I have a question. How accurate or how reliable is identification of IDUs from the records that we're talking about here for the denominator?

MS. BOSTROM: Sure. I think I'll have my colleague Diedra Gray can speak to the reliability of the denominator. Oh, for -- well, we don't have any testing data available. I think we can speak to -- I think that might jump into feasibility of it.

MEMBER TEUTSCH: Sort of completeness and, you know, accuracy.

MS. GRAY: Yes, so we did complete feasibility testing for this measure and all of

the required data elements were able to be captured in two different systems the feasibility testing was performed in and there were no issues with capturing the data. Am I answering your question?

MEMBER TEUTSCH: No, because I think it's really the sensitivity of the record for -- I would think specificity is probably pretty high. You don't catch a lot of people saying they're drug users who aren't. But I would imagine that there are a lot of drug users who you don't capture very easily. That's what I was thinking about how good is the denominator here?

MS. GRAY: So, I think I understand your question. So, like I said, all of the required data elements were able to be captured. The fact that you're focusing on the sensitivity of it and the differences in the medical records makes me think that perhaps it might be helpful to describe that there are different organizations and different vendors are able to create organizations specific and vendor specific

codes that are able to be mapped to the required data --- I'm sorry, to the standards that we use to specify the measure. So, in the case where someone wouldn't have been using the actual value sets that are required for reporting the measure they would be able to map to the required value sets even if they have an organization specific code or if they have a vendor specific code that's different.

Additionally, even though it's not applicable in this case, information and data can also be captured through free text in the EMR.

MEMBER TEUTSCH: What I'm really asking is, I'm sure that it can be coded and all that sort of stuff. The question is, how well is it recorded so that you actually have a true denominator?

MS. GRAY: Right. And so since we haven't performed the actual testing we wouldn't be able to answer that until after it's approved for trials ---

MEMBER TEUTSCH: Right. No, I get you

got the Hep C issues but in general, I mean it's a problem for IDUs I would think.

DR. WARD: I think you're right about the sensitivity specificity. I think that's one of the reasons we went to a demographic-based recommendation for the birth cohort in addition to its prevalence and other reasons. But I think that's -- and we've done studies to show that there's -- you miss a lot of persons who have this history because of the provider practices and desire and willingness to, you know, to explore that kind of history and the patient to divulge that kind of history.

So, it's a problem but I think at least for this measure, you know, we can at least track it for those that do have that history while we're seeking to improve that physician practice on our side of the issue here.

MS. MUNTHALI: Arjun.

MEMBER VENKATESH: Yes, I guess one is sort of related to that, would be more guidance as you put this into trial use and that is that

we have a very active screening and referral program in our emergency department for opioid overdose and so similarly I think since you're screening, you're trying to capture all the IV drug abuse, if you construct the measure to only capture that in the HR from a social history where that's captured you'll probably miss some of these. If you test this eventually to any site that you've got any data available outside of just the single practice notes I think it would be worthwhile and valuable information for us when we see this next time if you also screen for recent emergency department visit, hospitalization and all of their healthcare resource use associated with overdose, treatment, referral, things like that.

And then the other question I had that I think you're going to have to figure out on the specifications when you bring this back is what counts as the history of IV, active IV drug use because you're going to end up with in a one-year measure it's going to get a little tricky when

you've got somebody who was an active IV drug user also went to -- got screened at the end of last year, went to a detox program, came back out and so I think that that is probably going to be a big piece of the specifications to make sure there is no unintended consequences in screening.

MS. MUNTHALI: Okay. So, I think we're ready for a vote on 2b. Specifications. Consistent with evidence this is for eMeasure for approval for trial use. One is high, two is moderate, three is low and four is insufficient.

(Voting.)

MS. MUNTHALI: So, zero voted high and 10 voted moderate, three low so it passes, specifications.

And so we'll move to feasibility for Measure 3060.

MS. MUNTHALI: Arjun? Barry-Lewis? Jason?

MEMBER VENKATESH: I guess again, I think I would say that I do think at the data element level these are feasible. These are

things are captured in electronic health records. I think we'll have to see what the data looks like to actually assess how often things like a social history are being captured well and that that data is present across a wide variety of ages.

MEMBER HARRIS: I would concur that, you know, it's something that definitely could be captured.

MS. MUNTHALI: It looks like we're ready for a vote on feasibility. One high, two moderate, three low, four insufficient.

(Voting.)

MS. MUNTHALI: We're missing one vote and we got it. Zero high, 11 moderate and two low. So, for Measure 3060 it passes on feasibility.

And we'll move on to usability and use. Any discussion from Jason, Barry-Lewis, Arjun?

MEMBER VENKATESH: Nothing new. I mean, this is a tough one. It's hard to

interpret this when it's for trial use.

MS. MUNTHALI: Yes.

MEMBER VENKATESH: Because I feel like I should just say yes.

MS. MUNTHALI: Okay. So, one is high, two is moderate, three is low and four is insufficient.

(Voting.)

MS. MUNTHALI: So, zero voted high, eleven voted moderate and two low. So, this measure passes on usability and use.

And so for approval for trial use one yes, and two no.

So, eleven yes and two no, so Measure 3060 is recommended for approval for trial use.

So, we'll move on to the last measure for today -- we want to get a pulse check from you guys. Do you guys want to stop now and our developers call in from Chicago tomorrow or do you feel like we can push through this last measure?

CHAIR McINERNY: Push through.

MEMBER MOLINE: I think the value of having them there overwhelms our sense of getting out of our seats.

MS. MUNTHALI: Thank you, guys.

So, evidence really quickly. This is Measure 3061 and this is for those for follow up for patients identified with Hep C. Appropriate screening. So, I think we have with the exception of Arjun, Amy who is helping Jason and Barry-Lewis out with discussing this measure, so evidence.

MEMBER MINNICH: No pressure on the girl that gives the last one. Fortunately, I think it pretty much dovetails with the last two that were presented. The evidence is strong, although a bit dated to 2013 because of the new information that's coming forward. There were 30 observational studies that were reported and so I felt the evidence was strong.

MS. MUNTHALI: Any objections to that? Okay. I think we can move to a vote. One high, two moderate, three low and four insufficient.

(Voting.)

MS. MUNTHALI: Michael, I'm not sure if you submitted your vote. If you did can you submit it one last time?

MEMBER BAER: Did you get it?

MS. MUNTHALI: Would you mind sending it verbally, I hope you don't mind that.

MEMBER BAER: Moderate.

MS. MUNTHALI: Okay. I got it.

MEMBER BAER: Yes, I sent it twice.

MS. MUNTHALI: We just saw it, thank you.

MEMBER BAER: Okay. Thanks.

MS. MUNTHALI: Thank you.

Seven high, four moderate and two low so we'll move to performance gaps. Amy?

MEMBER MINNICH: So, because it is a new measure the developer just provided data from the literature. There were two studies reported that showed an obvious gap in performance. And also from the AASLD and IDSA that there was an estimated 13 to 18 percent of HCV infected

persons who have received treatment. There was also quite a bit of information relative to disparities looking at American Indians and Alaskan Natives that there is the highest incidence of HCV and poorest follow up. And that also African Americans, although comprise 12 percent of the United States population, they also have 22 percent of active HCV cases.

Minorities clearly show lower treatment rates and so there is a high degree of opportunity for improvement.

MS. MUNTHALI: Thank you. Anything to add? Okay. So, we'll vote on performance gap. One high, two moderate, three low and four insufficient.

(Voting.)

MS. MUNTHALI: Seven high, five moderate and one low for performance gap for Measure 3061. So, we'll move on to specification and we'll ask Amy to tee us up again.

MEMBER MINNICH: Sure. And in this regard since it is a trial measure it's just

looking at the reference to specificity.

MS. MUNTHALI: Steve?

MEMBER TEUTSCH: Because this is really talking about either people getting treatment or referral but there can be a big gap obviously between the referral and actually fulfilling that referral and then getting treated. So, I would really like to see some assessment of adequacy. Because follow up is always the horrendous problem on these major public health initiatives. Make sure that things that need to happen actually do happen. So, to the extent that you guys can assess that whether they actually got into care and got treated would be helpful to me because I think referral alone is probably only modestly effective.

DR. WARD: Of course, the referral has to happen for the other stuff to happen but we are very interested in monitoring the care cascade it's called from testing to cure and so we're accessing large data bases from health systems as well as from CMS. We also have a

cohort study in four sites looking at linkages to care at least in those settings. And even using this commercial lab data as you can track people from antibody, PCR, genotype to borrow clearance and at least three states where we're starting to collect all HCV data, positive or negative, so they can monitor that care cascade within their states by name. Massachusetts, New York, Kentucky and Tennessee. So, maybe that will become a trend in other states.

MEMBER TEUTSCH: Yes, that would be helpful. I mean, if you could show that referral that people think this is a high enough risk that they actually -- that's good enough and they actually do follow through then you don't have to monitor all that stuff all the time but my guess is you will.

DR. WARD: In the immediate future, yes, absolutely.

DR. NISHIMI: So, what I think I heard just so that you're clear here is, what I think the committee will be looking for when you come

back, this is your construct is treatment or are referred and I think they would like to see treatment and referred in two different bins so that they can see them to the extent that you can address that issue during your testing since they have the opportunity to comment on it now that would be very advisable.

DR. WARD: I think we wrote it that way because treatment is getting simpler so the tester can become the treater and maybe the referral is not as necessary. And so a primary care person can do both with the new drugs which were less of an option with the older ones.

MEMBER TEUTSCH: You know, when you look at referred, to me there's two parts. Did the primary physician refer the patient, number one. But number two did the patient go and see the specialist to whom he or she was referred? And, you know, obviously that second one is very important. And, unfortunately, sometimes we don't know and, you know, that goes for any referral across the board and I think more often

-- well, maybe things are getting better with electronic health records and communication between specialists and the primary care docs but in the past there's been a lot of referrals that never saw the specialist and that's a big problem.

MS. MUNTHALI: Barry-Lewis.

MEMBER HARRIS: So, I don't think that it's too far in the past just recently coming out as the chief medical officer for FQHC that part of our measures that we looked at in the FQHC world was whether or not referrals were completed for several different items of care. And so I think that there is the ability to capture data that shows whether or not their referral was completed because that's definitely things that are being looked at now because there's a lot of referrals happening and then they're not completed, as in the language --- I'm not sure if it was this one or the other one where it shows generally five percent actually got the treatment I think I read that right. So, I think that

would be important and would be something that -- I forgot your name, I'm so sorry.

DR. NISHIMI: Robyn.

MEMBER HARRIS: Robyn mentioned that we would be looking at coming back for the future, you know, whether or not they received -- they actually follow through even though I as family medicine would have referred but did they actually go to gastro or did they follow my treatment that I would have done?

CHAIR McINERNY: Thank you.

MS. MUNTHALI: No other questions on specification so we can vote. One high, two moderate, three low and four insufficient.

(Voting.)

MS. MUNTHALI: So, two voted high, nine voted moderate and two voted low so it passes on specifications that are consistent with the evidence. And so we'll move on to feasibility. Amy?

MEMBER MINNICH: And so two points are in the feasibility section. One is that the

measure specifications were consistent with the evidence. And, secondly, as a committee to look at the denominator exceptions. There were several that were listed under Threats to Validity and we're challenged to look and see if they were appropriately consistent with the evidence and I believe the answer is yes.

MS. MUNTHALI: Other comments? Concerns? Okay. It doesn't look like there are any so we'll vote on feasibility. One high, two moderate, three low and insufficient.

(Voting.)

MS. MUNTHALI: Looking for one more vote. Try again. Okay. So, feasibility for Measure 3061 one high, two moderate, one low. So, the measure passes on feasibility. Oh, two high, sorry, ten moderate and one low.

So, usability and use. Amy?

MEMBER MINNICH: So, from a usability standpoint the public health issue there is a high opportunity for improvement. Benefits do outweigh consequences and the planned use is for

quality improvement and bench marking purposes.

MS. MUNTHALI: Okay. No other questions or comments so for usability and use, one high, two moderate, three low and four insufficient. So, we're queuing up the slide.

CHAIR McINERNY: You have the wrong vote up. There it is.

MS. MUNTHALI: Okay. So, we're ready now. We'll try again.

(Voting.)

MS. MUNTHALI: Looking for two more votes. You can try again, two more votes.

CHAIR McINERNY: I have a question. Is there a receiver located in one place and are these things directional?

MS. MUNTHALI: No, only there, over here. Over here.

CHAIR McINERNY: Where is a receiver?

MS. MUNTHALI: Right here.

CHAIR McINERNY: Oh, so waving that way doesn't help?

MS. MUNTHALI: No.

CHAIR McINERNY: I've been in the wrong direction all day.

MS. MUNTHALI: So, one more vote if you could try again, sorry about that.

CHAIR McINERNY: There we go.

MS. MUNTHALI: Okay. One high, ten moderate, and two low so for Measure 3061 it passes usability and use.

And now we'll assess whether or not we should approve it for trial use and it's one yes and two no.

(Voting.)

MS. MUNTHALI: So, for Measure 3061, eleven have said yes to approval for trial use and two have said no for approval for trial use. Thank you all. Thank you to the Committee. Thank you, developers for your patience. And I'll turn it over to Tom.

CHAIR McINERNY: Well, I think one thing we proved is that the closer the dinner hour gets the faster we vote.

Oh, it's time for public comment by

the way. Public comment. Any public comment?

MS. MUNTHALI: So, Operator if you can open up the lines and then we'll see if anyone in the room has comments.

OPERATOR: If you would like to make a public comment please press star 1. And there are no public comments at this time.

CHAIR McINERNY: They've already left for dinner.

MS. MUNTHALI: Great. So, we are done for today. Tomorrow we reconvene for breakfast at 7:30. Sorry it's early. We have the majority of our new measures and we need to get through them. We're trying to get you guys out of here before the time on the agenda. And breakfast will be at 8:00, and again, reservations ---

CHAIR McINERNY: No, breakfast is 7:30.

MS. MUNTHALI: Sorry, the meeting will start at 8:00. Breakfast is at 7:30. So, half an hour earlier.

DR. NISHIMI: Can I just get a show of -- so the meeting starts at 8:00, breakfast is

at 7:30. Can I get a show of hands again to see who is going to dinner? Okay. And then so the restaurant is Georgia Brown's. It is at 950 15th Street. NQF is at 1030 so it's a block down 15th right across from the park there. So, heading in the direction of the park it's right across the park.

CHAIR McINERNY: 950 16th?

MS. MUNTHALI: 15th.

CHAIR McINERNY: 15th.

(Whereupon, the above-entitled matter went off the record at 5:31 p.m.)

E

X

Y

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Health and Well-Being
Standing Committee 2015-2017

Before: NQF

Date: 09-12-16

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter

NATIONAL QUALITY FORUM

+ + + + +

HEALTH AND WELL-BEING
STANDING COMMITTEE 2015-2017

+ + + + +

TUESDAY
SEPTEMBER 13, 2016

+ + + + +

The Standing Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:00 a.m., Thomas McInerny and Amir Qaseem, Co-Chairs, presiding.

PRESENT:
THOMAS McINERNY, MD, Co-Chair; Professor of Pediatrics, American Academy of Pediatrics
AMIR QASEEM, MD, PhD, MHA, Co-Chair; Director, American College of Physicians
JOHN AUERBACH, MBA, Senior Policy Advisor, Centers for Disease Control and Prevention*
MICHAEL BAER, MD, Network Medical Director, AmeriHealth Caritas Family of Companies*
RON BIALEK, MPP, CQIA, President, Public Health Foundation
JUAN EMILIO CARRILLO, MD, MPH, Vice President, Community Health, New York-Presbyterian; Associate Professor, Weill Cornell Medical College
BARRY-LEWIS HARRIS, II, MD, Chief Executive Officer, Common Table Health Alliance
CATHERINE HILL, DNP, APRN, Chief Nursing Officer/Director of Quality, Texas Health Resources

PATRICIA McKANE, DVM, MPH, Epidemiologist/SSDI Coordinator, Michigan Department of Community Health
AMY MINNICH, RN, MHSA, Director, Geisinger Health System
JACQUELINE MOLINE, MD, MSc, Vice President and Chair, Department of Occupational Medicine, Epidemiology and Prevention, Northwell Health (formerly North Shore-LIJ Health System)
MARCEL SALIVE, MD, MPH, Medical Officer, National Institute on Aging
KATIE SELLERS, DrPH, CPH, Chief Science and Strategy Officer, Association of State and Territorial Health Officials
JASON SPANGLER, MD, MPH, FACPM, Executive Director, Medical Policy, Amgen, Inc.
MATT STIEFEL, MPA, MS, Senior Director, Center for Population Health, Care Management Institute, Kaiser Permanente
STEVEN TEUTSCH, MD, MPH, Adjunct Professor, Fielding School of Public Health, University of California, Los Angeles; Senior Fellow, Public Health Institute; Senior Fellow, Schaeffer Center, University of Southern California
ARJUN VENKATESH, MD, MBA, Robert Wood Johnson Foundation Clinical Scholar, Yale University School of Medicine

NQF STAFF:
ELISA MUNTHALI, MPH, Vice President, Quality Management
SHEILA CRAWFORD, Administrative Manager
DIANE FERGUSON, Administrative Assistant
ROBYN NISHIMI, PhD, Consultant
YETUNDE OGUNGBEMI, Project Analyst

ALSO PRESENT:
DALE BRATZLER, DO, MPH, Co-Chair, NQF Pulmonary and Critical Care Standing Committee*
DANIEL H. GREEN, MD, FACOG, Centers for Medicare & Medicaid Services*
IRENE HALL, PhD, MPH, Centers for Disease Control and Prevention*
JOE LYNCH, Avalere Health
SHARON McCAULEY, MS, MBA, RDN, LDN, FADA, FAND, Academy of Nutrition & Dietetics
MEREDITH PONDER, JD, Defeat Malnutrition Today*
RAMESH C. SACHDEVA, MD, PhD, MBA, JD, Children's Hospital of Wisconsin*
ANITA SOMPLASKY, RN, Quality Insights of Pennsylvania
ALISON STEIBER, PhD, MS, Academy of Nutrition & Dietetics*
CAROL STOCKS, RN, MHSA, Agency for Healthcare Research and Quality
ANGEL VALLADARES, MPH, Avalere Health
ABIGAIL VIALL, Centers for Disease Control and Prevention
KERIANN WELLS, MPP, Mathematica Policy Research
DONNA M. WOODS, PhD, EdM, Northwestern University

*** present by teleconference**

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

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

(8:07 a.m.)

MS. MUNTHALI: Good morning everyone. I think we're going to get started.

Shan, is the public line open?

OPERATOR: Yes, ma'am, you are live.

MS. MUNTHALI: Okay.

Thank you everyone. Thank you for coming back for Day 2 of the Health and Well-Being Standing Committee's In-Person meeting.

My name is Elisa Munthali and I'm joined here by my colleagues and our Co-Chair, Tom McInerny and Amir Qaseem will be joining us, he's the other Co-Chair, just after 9:00.

We just wanted to very briefly go over -- yesterday was a great day. We reviewed 14 measures, which is quite a lot for an NQF committee. And, all but 12 passed with, you know, regular NQF endorsement recommendation.

One was recommended with reserve status and one consensus was not reached.

And, as we mentioned yesterday, we'll

be resolving the consensus not reached issues during the post-comment call.

And, so, today, we have ten measures that are under review. All but one of them are new measures. So, we're very happy about that, the potential of new measures coming into the NQF portfolio.

So, we'll start with the first one.

Tom, I don't know if you have any remarks before we start, sorry.

CO-CHAIR McINERNY: Well, again, I just want to thank everybody and the staff for a great day yesterday and all of their help as we work through a large number of measures.

And, I'm so glad that we now have flu vaccine out of the way. I think we've had it up to our eyeballs with flu vaccine.

And, I think today, we have some very interesting new measures that'll be -- will need some very good discussion.

And, to me, one of the ones that's the most interesting is the first one that we're

about ready to consider. As a pediatrician, it's kind of near and dear to my heart. So, let's see how this one goes, please.

Thanks so very much.

MS. MUNTHALI: Thank you, Tom.

So, the first measure is Measure 3071, it's Follow-up Referral After Positive Developmental Screening. It's developed and steward by Northwestern University.

And, I'm not sure if Northwestern is here in the room.

And, so, as we've been doing with the meeting yesterday, a two to three minute intro and then we'll turn it over to our lead discussants.

DR. WOODS: First off, I want to ask if any of the rest of the team is on the phone so that they can contribute.

MEMBER AUERBACH: John Auerbach is on.

DR. WOODS: Dr. Sachdeva, Dr. Tate?

MEMBER AUERBACH: This is John Auerbach. I'm not sure if they'll --

DR. WOODS: Hi, John.

MS. MUNTHALI: He's one of our committee members. Welcome. Actually, John, while you're on the phone, if you can just introduce yourself before we get into discussion and let us know if there is anything you'd like to disclose.

MEMBER AUERBACH: Hello, everybody. My name is John Auerbach. I'm the Policy Director and Acting Director of State, Local, Tribal and Territorial Support at the Centers for Disease Control and I have no conflict to announce.

MS. MUNTHALI: Thank you.

I'll turn it over to you to see if your colleagues are on the phone.

DR. WOODS: Okay. My name is Donna Woods. I'm faculty at Feinberg School of Medicine at Northwestern in the Department of Pediatrics and the Center for Healthcare Studies.

So, developmental screening and follow up are fundamental aspects of pediatric care,

albeit, very challenging.

Thirteen percent of children in the U.S. have developmental or behavioral disabilities and fewer than half of children with delays are identified prior to starting school.

When a delaying diagnosis of treatment occurs, critical, often time sensitive early brain and child development opportunities are missed.

Currently, in the literature, we understand that 34 to 30 percent, 7 percent, of high risk infants and 61 percent of young children who fail a developmental screen are not referred for any further evaluation or treatment. This is a considerable performance gap.

We tested this measure as a chart review measure. The data elements are generated by the healthcare personnel and then the data elements are extracted from the record.

We tested this in four institutions in the Chicago Pediatric Quality Safety Consortium which consists of a safety net hospital, a

freestanding children's hospital.

The primary care networks for this safety net hospital is freestanding children's hospital and two suburban children's hospitals networks as well as Ashe Pediatrics in Pennsylvania.

In the testing within the Chicago Pediatric Quality Safety Consortium, performance varied from 31 percent to 100 percent with performance scores of 31 percent, 40 percent and a 100 percent.

And, Ashe Pediatrics performance was 23 percent.

Reliability results, the use of a validated tool was 93.6 percent agreement of patients, with positive developmental screening results had agreement of 99 percent and patients who received a referral for follow up care within seven calendar days of receiving a positive developmental screen had the agreement of 73 percent.

We tested -- so, that's the results of

our testing.

And, we also had a stakeholder panel. It included both patient family advocates as well as pediatrician school personnel, early childhood daycare personnel, nurses, physicians, developmental physicians, neurologists, et cetera.

All agreed that this was an important measure, that this was consistent with the guideline recommended care for pediatric patients.

There's a 2006 guideline that was reaffirmed in 2014, which is the basis for Bright Futures Guidelines and is built on systematic reviews.

MS. MUNTHALI: Thank you.

So, our lead discussants are Katie, Tom and John. And, I know, John, you have limited time with us, so, perhaps you'd like to start off the discussion on evidence?

MEMBER AUERBACH: Let me just -- I'd rather not start only because I didn't join you

yesterday. So, maybe someone else can start.

MS. MUNTHALI: Okay.

MEMBER AUERBACH: And, I will happily join in.

MS. MUNTHALI: I think Tom's ready.

CO-CHAIR McINERNY: No, no, Katie just twisted my arm and said she wanted to go first.

MEMBER SELLERS: I'm not sure that's exactly how that conversation went. No, I'm happy to.

Just, overall, you know, this seems to me to be a very important issue. I think the developers certainly made their case on that.

They don't cite evidence regarding the effectiveness of referrals, per se. But, this is kind of similar to a discussion we had yesterday, I think it was regarding Hep C. I mean, certainly, the referral is the next step to getting the needed care. And, they do cite evidence regarding the effectiveness of early intervention.

I think the issue with the evidence

here is it doesn't quite make it through the algorithm because the evidence is not specifically on the referral, per se.

So, when the staff put it through the algorithm, it came out as insufficient with exception. I think the exception being, you know, can we -- knowing that there -- knowing that we don't have evidence on the effectiveness of referral, per se, but, we do have evidence showing that intervention is important.

And, we can say from common sense that that referral is a way to get to that intervention. Can we do insufficient with exception rating, thereby, holding providers accountable for doing something without empirical evidence?

I'm not sure if I'm articulating that clearly. But, that --

MS. MUNTHALI: You are. So, maybe we can pull up the algorithm, it might be helpful.

Sheila, can you pull up the evidence algorithm?

MS. OGUNGBEMI: It's on our committee home page if you can just go to the general documents at top.

DR. WOODS: Also, if I could just speak.

Dr. Sachdeva has now joined the call, a member of our team. So, if there are questions for him as well.

Dr. Ramesh Sachdeva, who, at the time, was the VP for Quality Informatics at the AAP who is the PI for the Pediatric Measure Center of Excellence.

CO-CHAIR McINERNY: Hi, Ramesh, it's Tom McInerny, how are you doing?

DR. SACHDEVA: Good morning. How are you? And, I believe Dr. Tate may be joining us, but I don't have confirmation of that. And glad Donna and I be a part of the team that developed this measure.

CO-CHAIR McINERNY: So, I'll put in my two cents at this point.

You know, this, to me, is a tough

measure and I probably have a little bit of a conflict of interest since I'm past President of the American Academy of Pediatrics.

But, you know, we do have Measure -- I think we approved Measure, a while back, 1448 which says that pediatricians and other clinicians who care for children should be doing developmental screening. So, that step is clear and that should take place.

We also have evidence that early intervention for children with developmental delays improves the outcome.

Unfortunately, what we do not have really good evidence, and this is where the U.S. Preventative Services Task Force weighed in, is the step between the screening and when the patient screens positive and the referral. Does the referral actually lead to an improved outcome?

Now, common sense would say, yes, it does. And, certainly, the American Academy of Pediatrics feels strongly that it does.

But, then, we have that darn USP -- Preventative Services Task Force which, and occasionally, has bothered me in the past about some other measures.

But, they really look at things, at evidence, very carefully and very closely. And, I think we do need to pay attention to that.

The other point I would like to make is that I'm not comfortable with just -- I think what we're saying is that, in the chart, the patient -- the clinician says patient did not pass the screening test. And, there are several validated, well validated screening tests that are being used as standardized screening tests.

And, then, says, patient referred to XYZ for follow up. To me, what really we need to know is, did the patient actually go to that -- or say, for further testing at that center?

And, I would like to see that there would be somehow in the chart a notation that, in fact, the patient was evaluated. And, then, I think that would make this a much stronger

statement, a much stronger measure.

DR. WOODS: If I could respond to that? This is one of a set of measures on all of the follow up that is appropriate for the care of a pediatric patient with a positive -- with a developmental screen.

One is to have the communication with the family. Afterwards, regardless of the result, this measure, which is about the referral, because, as you can see, referral is, unfortunately, very -- is not happening for 37 percent and 61 percent of older children.

The third measure, which is currently undergoing further testing is exactly what you said, which is the referral tracking and follow up to make sure that the family actually did follow up and that there has been engaged active further evaluation or treatment.

CO-CHAIR McINERNY: John, do you have any comments about the evidence?

MEMBER AUERBACH: Thanks, thanks for asking.

I guess I have a couple of questions maybe related to evidence. And, I do apologize that the other people feel like these are self-evident.

But, there's some of the questions that I had and just would love to have the developer respond it.

You know, the first one is just the seven-day time limit. You know, is there a particular reason for seven day other than that just seems like a reasonable amount?

The second question that I have is, is there evidence about, and again, I do apologize when everybody else would say that this is self-evident, is there clinical evidence that a referral is always the appropriate clinical next step as contrasted with, say, the pediatrician scheduling a follow up visit with the patient or monitoring the development over time?

In part, I say that because I remember, at times, talking to pediatricians who would say the -- it wasn't clear enough, whether

or not this was appropriate for referral and maybe let me monitor it for a little bit longer and have the patient come back, the family come back next month or so?

And, then, the third question I guess I have is the evidence with regard to the availability of appropriate referral services if there's evidence about this ready access to those services such that we could have some level of comfort that a referral would be realistic or likely to occur?

DR. WOODS: I can respond to those questions. The seven day determination, there was generally believed and thought that a positive developmental screen should receive a referral on the day of the screen.

But, there was a recognition in the nature of practice that that was not always possible. So, the expert work group determined that it was -- it would be overly potentially chastising clinicians who had all of the best intentions and all of the best practice that it

wouldn't impede the care of a patient if, for some reason, it took a week to get the referral in.

So, there's no harm in waiting seven days, however, most should occur within the actual visit period, that actual screening visit.

So, that's the answer to your first question. The second question was -- one was about watching and waiting. I think that was the third question. And --

MEMBER AUERBACH: Yes, the second question was about -- yes, the second question was about watching and waiting and/or scheduling a follow up appointment with the primary care.

DR. WOODS: Correct.

It was believed that frequently, inappropriate practice, based on the guidelines, based on the evidence, that it was inappropriate practice to just wait until the next evaluation appointment.

If the clinician felt that there might be a reason to wait, it should not go more than a

month, that they shouldn't wait through the entire next developmental screen window. Because, as we said, there can be harm due to the time sensitive nature of many developmental delays.

And, you had one further question.

MEMBER AUERBACH: The availability of referral services.

DR. WOODS: So, the requirement is to make a referral. It is not to receive the services of the referral, but it is to make the referral.

And, the idea is, again, as soon as possible. And, as you're probably aware, that is particularly in some urban environments, there can be a backlog. So, reaching out as soon as possible is very important.

So, these are responses based on expert panel review and determination.

MS. MUNTHALI: Other comments or questions on evidence?

Steve?

MEMBER TEUTSCH: I have two questions.

One, John sort of got at and hopefully you can elaborate. Aren't there some of these that the pediatrician can manage themselves so that they don't naturally require a referral? They may have resources within the practice.

The other question I had is, this is about referral once kids are detected. Can you say something about how often they're detected using these more formal instruments? I mean, how many are we -- I mean, that's about the denominator for this, right?

So, it wasn't clear to me whether there isn't a big gap in just finding them to begin with.

DR. WOODS: So, to your first question, if a provider refers within their practice to follow up services, that counts as a referral. And, we actually documented in Ashe Pediatrics how that can occur because they had additional resources within that practice.

So, yes, and, it is counted because

it's just, you know, the referral is described. Right? So, it becomes clear what the referral is, it doesn't have to go outside of your practice if your practice actually has those services.

In terms of what we know about developmental screening with validated tools, the research and the work from Bright Futures, when we started this project, had the rate of children being evaluated using validated tools in the range of 60 percent. And, across the period, it went up into the 80s across the country.

However, in our study, we discovered a fairly disturbing disparity in that the safety net providers in our sample, again, only used validated tools about 38 percent of the time.

So, when a child is evaluated by, just like, oh, he looks okay or, oh, she looks okay, as opposed to really doing a systematic review, a lot of potential children can be missed.

So, we also, as you saw in the denominator, it requires that a validated tool be

used to detect a positive screen. Because, otherwise, you can't know that the referral is required.

CO-CHAIR McINERNY: Steve, is there a metric? And, do we have a measure, then, that asks about whether they were screened?

DR. WOODS: Yes.

CO-CHAIR McINERNY: Or using --

DR. WOODS: It's already endorsed. It's one of the initial core measures, and so, we're building on that.

CO-CHAIR McINERNY: 1448. And, my feeling is that it's about 20 percent of patients who are screened using validated instruments fail.

And, then, I think the question is sort of a self-referral. You know, in other words, I say I'm a pediatrician who's interested in developmental problems and I can handle this. I think that's sketchy. Yes, I think that can be difficult.

Most of the time, you need physical

therapists, occupational therapists, social workers, things like -- psychologists, et cetera. And, most pediatricians don't have those skills.

So, I think that would be a little -- I don't think I would accept that as a self-referral.

MS. MUNTHALI: Arjun and then Jason.

MEMBER VENKATESH: So, I guess, on the evidence question, the question I have is, and this is, I think, challenging because, the measure itself is really, really narrow. Right?

The measure is about whether or not there's documentation of a referral. It's not all these other things that we want the measure to be that you mentioned, which is the actual completion of the referral visit, something along those lines.

And, so, I get that there's not going to be any literature or research on referrals. It's a highly unstudied topic.

Is there at least some qualitative work that would suggest that that is the barrier

to the follow up? Like something from the patient side that would say, hey, lack of referrals or not being referred is the reason I didn't see somebody.

Because, then at least you could say, hey, these referrals, even if there's no evidence for them, because they're never going to be studied, have some notion of potential value.

DR. WOODS: Like I just reported, 34 to 37 percent of high risk infants and 61 percent of young children who fail a developmental screen are not referred for further evaluation.

So, at that point, we know there are a lot of children who aren't getting the services that are required and, actually, honestly, really expected from pediatric care.

There's -- from the beginning work we've done on the next measure, which is on referral tracking, let me see if I can find the actual, because I brought it with me just in case, practices that successfully track referrals have found that some families did not follow

through with referrals.

Tracking referrals led to better communication with local referral sources and more children were identified and linked to services.

So, it is the combination of -- they're not -- the children are not going to get there if they're not referred.

And, then, if the referral is not tracked, appropriate follow up care cannot occur, making sure that the family follows up as well as understanding the nature of the further evaluation or treatment that is occurring for that pediatrician's patient.

But, one of the things that we found in the beginning -- the initial testing of the referral tracking measure is so little referral tracking is going on.

So, we started with the measures of patients who are referred and we ended up with seven patients in the Chicago Pediatric Quality and Safety Consortium and four in Ashe Pediatrics

from over 200.

So, you know, we're finding that referral tracking is not occurring either. So, we're having to reselect a sample so that we can start with that referral point.

MS. MUNTHALI: Jason and then Tom and the Matt.

MEMBER SPANGLER: I want to go back to something Steve mentioned about the measure for actual screening.

We're talking about 1448. That was a measure that was supposed to go under maintenance two years ago, didn't. It's time limited. Do we know the status of that and what --

Because I know we were waiting for -- I'm looking at a report from last year for that. And, I'm just wondering where that is because is that going to disappear?

Because, I think that has direct relevance even to this measure because, don't they have a -- there had to be testing done and analysis and --

DR. NISHIMI: Yes, Jason, you're correct.

That measure, they weren't prepared to bring forward information at this time. So, we gave them another deferral.

MEMBER SPANGLER: For how long? Because the last time it was endorsed was 2011.

DR. NISHIMI: Right.

MEMBER SPANGLER: And, so, we're talking five years now. And, there were two measures that were very similar and I know NCQA pulled their measure which --

DR. NISHIMI: They pulled their measure.

MEMBER SPANGLER: -- was almost the same.

DR. NISHIMI: Right.

MEMBER SPANGLER: So, I'm just -- because, that's going to affect this -- I mean, there's a relationship between both these measures.

DR. NISHIMI: It's unclear whether the

developer will be able to obtain sufficient information even by the next cycle. I'm trying to be politic about this, but candid.

DR. WOODS: We didn't develop that measure, but we built that measure into our measure. We thought that it has to be the basis. I could provide further information, if that would be helpful to this group following this meeting on what we found regarding the use of a validated tool and additional updates on the --

Because there was a reaffirmation of the use of validated tools by both the American Academy of Pediatrics Guideline and also the Bright Futures.

MEMBER SPANGLER: Well, my --

DR. WOODS: So, the updates are --

MEMBER SPANGLER: My concern is not that -- my concern is, I think that's a good measure. My concern is, if you have that measure built into your measure and that measure disappears, that affects your measure.

DR. WOODS: It doesn't affect the

guideline, though.

MEMBER STIEFEL: Maybe dwelling on evidence is appropriate in this one because it's the challenge. And, it seems like a flaw in our process, plucking out a step in a process and evaluating the evidence of that step in the process, I guess, following up from what Amir was saying, it's impossible to demonstrate the evidence of that one step in the process.

I'm talking about the screening, referral and referral follow up and plucking out referral and evaluating the evidence for referral is not really -- nor consistent with the sort of health and well-being focus of our work.

This step, this measure of a step in a process does not contribute to health and well-being, the whole process does.

So, maybe it's just a question about how we handle an isolated measure in a process where the whole process has to take place to improve health and well-being.

DR. WOODS: If I could speak to --

DR. NISHIMI: No.

DR. WOODS: Okay, sorry.

DR. NISHIMI: We look at lots of process measures that involve several steps. And, so, it's for the committee to decide whether the step, you know, more distal from the initiation point is the appropriate place to measure and hold providers accountable for.

Or, and, there are sometimes is evidence in the series of steps, you know, related to the intervening events. In this case, there isn't for the referral step. So, it may be that the committee, again, votes for evidence with exception. That's why that's an option for you because you feel it is, you know, one piece and it's okay to let it move forward because there isn't that one step.

Or, you may decide to wait for, you know, the additional measure that's being developed on actual referral. Or, you can, you know, turn it down in its entirety.

CO-CHAIR McINERNY: Well,

interestingly, in the latest issue of AAP news, they do discuss the AAP, it says AAP stands by recommendations on universal developmental screening.

And, they say, when a development disorder is suspected, the pediatrician should simultaneously begin the diagnostic process, refer to a specialist for final diagnosis and then refer for therapy.

For this process, it is hoped the diagnosis and treatment can be instituted earlier with improved outcomes.

So, you know, again, this is more of an expert opinion. There was a policy statement in the AAP made in 2014 that elaborated this. And, certainly, it's sort of the dogma of the AAP and it is in the Bright Futures.

For some of you who may not know, Bright Futures is essentially the guideline for how to do anticipatory guidance, screening, et cetera for children and it has been recognized by the CDC and the federal government. It's part of

the Affordable Care Act so that all of the recommended procedures, it's in the Bright Futures to be covered at no charge to the patient. That's how strongly it's felt that this is a good guideline.

But, again, I think the crux of the matter is that it is based on expert opinion and there is no Grade A kind of evidence by testing, et cetera.

DR. NISHIMI: Amir? Amir, I'm sorry, Amir? Arjun?

MEMBER VENKATESH: I was just wondering if you guys can -- no problem -- I was wondering if you can scroll on the chart up here to the insufficient evidence page, the next page?

And, the reason I think is because what Tom just said, it sounds to me like, for this measure, we are absolutely kind of in Box 10 and trying to decide on 10 and 11.

Which is that, when you don't have actual empirical evidence, you move off the front page and Box 11 is what allows you to say there's

insufficient evidence with exception if there is a national or international consensus recommendation, which is what it sounds like the AAP recommendation is.

The question you have to ask before you get to Box 11 is Box 10, and there's an example measure that is sort of similar in construct. But, I think they give the examples, there's a proposal to measure whether blood pressure is assessed each visit instead of blood pressure control.

Here, we're trying to measure whether or not a referral is documented, whether or not -- rather than whether or not a referral happened.

And, so, the thing I'm a little unclear about is so, if you think that there is another measure already available that does the like blood pressure control, then you would say, yes, and go down and say insufficient is the way I'm reading this.

And, so, I guess there's no other measure of right now, anything downstream of the

initial screening measure that we reviewed before.

And, so, I guess it's no, and then, I would go down that path to rate as insufficient evidence.

But, then, I heard that there's a proposal that said, hey, we're working on a measure that's going to be the actual referral.

And, so, if you -- if I think about that measure, that they're actually developing a measure about the absolute referral as a committee and if somebody on this, I'd want to say, hey, that's the measure we want. That's a real process measure.

This is more of a kind of checkbox or, you know, yes, I documented that there was a referral type measure which is a much lower value.

MS. MUNTHALI: So, you should be looking at the measure that's in front of you. I know there are aspirations for Northwestern to develop a measure that's probably going to

capture a lot more. But, they've brought forward this measure.

So, on the merits of this measure and, Arjun, you're very right about how you go through this algorithm.

And, if you -- if the committee agrees as Arjun has said, that, you know, there's an exception to this, then that's the way you should go.

But, it is on the merits of this measure.

DR. WOODS: Yes, and it is in the initial expectation as a parent, even, we have many parent and family advocates on our group as well as clinicians that their expectation is that, if there's a positive developmental screening, there will be some care for their child. And, they advocate for that.

And, it's not the same as blood pressure assessed and blood pressure controlled. Because there is a clear, you know, a clear pathway of treatment.

DR. NISHIMI: Is there anything else from the committee? Arjun, was your --

Is the committee ready to vote on evidence?

Okay, just as a reminder, if we want to consider the orange path here, then to invoke the insufficient with exception, on the first vote overall on evidence, which will be the, you know, high, moderate, low, insufficient vote, the measure needs -- you need to achieve consensus to go down the insufficient route.

And, then, we'll hold the second vote on insufficient with exception.

So, is that clear to folks? If it's low, if it fails because you think it's low, then we don't march down the insufficient.

Okay, go ahead.

MS. CRAWFORD: For Measure 3071, we're voting on evidence, one for high, two for moderate, three for low, four for insufficient.

We're looking for 16 votes. And, voting is open.

(Voting.)

We have 0 for high, 2 for moderate, 2 for low 11 for insufficient.

DR. NISHIMI: Okay, so, we'll vote on insufficient with exception. This is for Measure 3071.

MS. CRAWFORD: And, our choices are one for high -- no, one or two for Measure 3071.

DR. NISHIMI: So, one will be insufficient with exception and two is no exception. So, then, the measure would fail.

(Voting.)

MS. CRAWFORD: And, one more vote, please?

CO-CHAIR McINERNY: It's like Dancing with the Stars here.

MS. CRAWFORD: Ten for insufficient evidence with exception, 5 for no exception.

MEMBER SPANGLER: Sorry, this is a process question. Does everybody vote for this or only the people who voted insufficient?

MS. MUNTHALI: Everybody.

MEMBER SPANGLER: Okay.

MS. MUNTHALI: We'll proceed to performance gap. Our lead discussants, Tom, John or Katie?

MEMBER SELLERS: Sure, so, moving on to performance gap, right?

So, they did their own -- they presented their own evidence on this where they had a very small sample because, I don't -- I did not manage to pull the numbers up in front of me, but, they only wound up with like 16 positive screens in Chicago and maybe 12 in North Carolina.

But, they do summarize the literature which shows quite a gap. And, the developer has referred to this a few times in our conversation today.

So, it seems there's a gap somewhere in the neighborhood of 35 to -- sorry, only 35 to 61 percent of children are being referred after the positive screen.

So, they did not present any data on

disparities, but it does seem to me like a fairly high performance gap and plenty of room for improvement.

DR. NISHIMI: Any other comments?

Okay, let's vote on gap. Voting on gap for Measure 3071.

MS. CRAWFORD: One is high, two is moderate, three is low, four is insufficient.

Voting is open.

(Voting.)

Fifteen votes are in, 6 for high, 7 for moderate, 2 for low, 0 for insufficient.

DR. NISHIMI: Okay, so, we'll move on to reliability and validity.

MEMBER SELLERS: So, looking at reliability, I think there's a couple of issues here.

One has to do with the definition of referral and the other has to do with that small sample size they had for the testing.

So, when you look at the way referral is defined, basically, what they have, it says,

referral for follow up care is defined as the formal event by which the clinician provides a referral to the patient family. It does not include any further steps in the process like securing the appointment, et cetera.

And, refers the patient and family for further evaluation or to any type of therapy, intervention or education to mitigate developmental delays.

And, referral can be made within the medical home or outside the medical home. A referral can include the form of watchful waiting by which the clinician offers practice based interventions and schedules the follow up visit within three months.

Some referral types are listed below, but this list is not exhaustive. And, then there's maybe 12 different items listed there.

So, I think the question is, you know, is this likely to be coded in a reliable manner? And, unfortunately, the testing they provided, you know, because they only found 16 positive --

15 positive screens, it's a little bit hard to tell how reliable how this would be.

MEMBER VENKATESH: I guess my concern is the fact that they found so few. Doesn't that actually suggest that the reliability is lower or this is where the reliability and validity start to tie together a little bit because the presence of the CPT code is how it's being identified. It's not being currently coded.

And, so, I think that we're kind of in this weird box on the reliability one that is -- there's no empirical reliability testing.

And, then, you're sort of saying, okay, was empirical validity testing or patient level data conducted?

And, so, we're kind of going down that path, I think. But, I'm pretty sure that we can't, based on what's here, say that this is, you know, sufficient reliability testing.

DR. NISHIMI: So, yes, they did data element level validity testing which we bring forward for the reliability testing score.

So, you are correct, they did not do score level reliability testing. They did data element validity testing which NQF qualifies for under the reliability rubric.

DR. WOODS: What -- were these school level?

DR. NISHIMI: Signal to noise. You would have to have signal and noise and many more entities than this.

DR. WOODS: So, five is insufficient?

DR. NISHIMI: They're finding to be potential, but, I would be suspicious that you'd be able to do it off of five.

Any other questions or comments on the reliability?

Ready to vote?

MS. CRAWFORD: We're voting on reliability for measure 3071. One is high, two is moderate, three is low, four is insufficient.

(Voting.)

Voting is closed.

Zero for high, 6 for moderate, 7 low,

2 insufficient.

Forty percent moderate, 47 percent low and 13 percent insufficient.

MS. MUNTHALI: So consensus stands.

DR. NISHIMI: Consensus not reached on reliability, so we'll proceed to discuss the validity criterion.

Right, so, 40 to 60 percent, it's inclusive in that range. So, if it had been 39 percent moderate, then it would have failed. But, 40 percent is consensus not reached.

So, for validity, Katie?

MEMBER SELLERS: Yes, so, with the validity, there are a couple of issues here.

One has to do with that specification of seven days. We did hear the rationale for that, which made sense to me.

But, on the validity testing, they did it at the level of the measure score and it was face validity only. And, it was done through an open comment period by stakeholders.

And, the developer reported more than

100 individuals commenting. And, that 65 percent of the respondents agreed that the measure is extremely valid.

So, I guess I would like to hear a little bit more about what questions were posed and what the qualifications of the stakeholders reporting on this are?

DR. WOODS: We used our broad expert panel network to reach out across the country to a variety of provider associations, to educators, to policy bodies, to patient and family stakeholder and advocacy organizations to make sure we have the broadest review.

And, I think we also wanted to, through this process, notify all relevant stakeholders that this was something that was coming down the pike.

DR. NISHIMI: Any other questions or comments on validity?

MEMBER AUERBACH: When there were questions raised, what was the nature of the questions?

DR. WOODS: Oh, gosh, I didn't review that. There were some questions about the watchful waiting, I remember that and what was going to be sufficient for watchful waiting.

There were questions about whether a child who already had a positive developmental screen should get a regularly scheduled developmental screen.

There was strong advocacy from parent and family organizations that, if the parent -- that parents believe that their child, even with a positive developmental screen should have a developmental screen again to review where the child is and that to exclude them would be not good, healthful care and would impede the relationship between the pediatrician and the family.

There were two other things, but I'm not -- I'm -- I'll apologize to you right now, I know there were two other things. I can get back to the committee when I get home and review that file.

DR. NISHIMI: Arjun and then Ron.

MEMBER VENKATESH: I guess the only part about the validity testing, I am -- I totally get face validity and its role in many, many measures require it.

The only thing that throws me off a little bit is the question they ask, I think, is whether or not they thought the measure was valid. And, that's a tough term and a tough thing to answer because, we have debates here and like 100-page documents around what is and what isn't validity.

And, so, usually, I feel like when I see these face validity surveys, the question that's asked is along those lines is, do the specifications of this measure line up with evidence?

And, then, something about the linkage of this measure with an outcome. So, would this measure advance the quality of care or is the -- would a higher performance on this measure be associated with better outcomes for children who

screen positive? Something like that.

And, so, were there any of those kinds of questions in the survey that would just say, there's a linkage between this process and an outcome or that this recommendation lines up with guidelines?

I guess they sort of do because they say it's extremely important, but that's sort of loose.

DR. WOODS: Yes, there was a fair bit of comment from a wide variety of stakeholders that the way the measure was specified was an appropriate guideline-based, valid, feasible method for assessing what was determined to be a fundamental and critical aspect of pediatric care that is demonstrating very poor performance at this time while being a fundamental expectation of parents and families.

DR. NISHIMI: I'm just going to ask, before I go to Ron, NQF requires that for face validity at the measure level, you ask the group whether they think the measure, as specified, but

we'll say the measure is score -- the measure score can distinguish good from bad quality and is an, you know, indicator of quality of care. Did you ask about that?

Not whether they commented, did you ask that?

DR. WOODS: Yes, we did ask if they felt that this particular measure, as specified, could distinguish high versus poor quality performance.

DR. NISHIMI: Okay.

Ron?

MEMBER BIALEK: Do you have data on -- excuse me -- the proportion of respondents from suburban, urban, rural and frontier?

DR. WOODS: Yes.

MEMBER BIALEK: And, then, also, the differences in responses.

DR. WOODS: Nothing from frontier. But, suburban, urban and rural, yes.

Okay, let me just make sure I'm looking at the right --

Rural, 23 percent. We had from one suburban -- what -- performance met the measure.

MEMBER BIALEK: I'm sorry, in terms of asking the question about the validity.

DR. WOODS: Oh --

MEMBER BIALEK: So, I'm trying to get to when you reached to a variety of stakeholders, the -- how many of them, or the proportion, when you did this, how many were urban, rural, suburban? Sorry, yes, suburban? And, then, what the rural response was to the validity question that was just posed.

DR. WOODS: I did not do that analysis. I can go back and do that analysis and provide you with more information.

We were working with -- I mean, so, we had Head Start and early intervention programs represented in a variety of settings. But, I didn't say early intervention in rural versus early intervention in the city comments or patient advocacy in the rural versus patient advocacy in an urban context.

MEMBER BIALEK: Well, it comes down to the provider mix that may exist in rural versus urban and suburban. That's what I'm thinking about is the rural provider, you may have fewer choices and I didn't know if there was a difference in response for validity based upon the rural provider experience.

DR. WOODS: There was some little discussion about whether there was better access in rural versus urban environments. And, this was in the expert work group discussion of the results of the public comment.

And, there was quite a bit of controversy about whether the urban folks had potentially less access because there was greater numbers of children needing services. And, that the lengths of wait times were -- it was voiced sometimes longer.

DR. NISHIMI: Matt?

MEMBER STIEFEL: It seems like this discussion of validity rehashes the discussion of evidence and presents the same problem of

extracting a step in the process and asking the almost non-answerable question about, is this step in the process valid?

I suspect that there's not an exception in terms of our review of validity, but it seems like the same issue that we had with evidence.

DR. NISHIMI: Certainly, around the specifications. But, the other question is whether the testing was adequate, so, whether you feel that, you know, face validity was, you know, fine, the process that she described, whether all the threats to validity, which we haven't discussed, but I was going to raise with the committee, have been assessed and you're comfortable with.

So, whether their description of how they handle missing data. They could not -- the measure isn't risk adjusted, whether you think that's appropriate.

They could not demonstrate meaningful differences among measured entities because they

had too few in at the end of the day.

So, the question is not just about the validity of the specifications and the evidence underlying that, but also the testing for face validity, meaningful differences, risk adjustment or lack thereof and how they handle the same data.

Any other questions on the validity? Committee ready to vote?

MS. CRAWFORD: Voting on Measure 3071 validity, face validity only, one is moderate, two is low, three insufficient.

Okay. Voting has closed at 14. Results, 2 for moderate, 5 for low, 7 insufficient, 14 percent moderate, 36 percent low, 50 percent insufficient.

DR. NISHIMI: So, the measure does not pass the validity testing.

Does the committee have any additional comments that they'd like to convey to the developer beyond those that we've touched upon?

I think you heard that there's a lot

of interest in this measure. Perhaps if you could beef up that validity testing or NQF's happy to discuss that with you afterwards and maybe a little technical assistance on some construct validity or something like that would be helpful.

MEMBER TEUTSCH: I think to get back to Matt's comment, I think if we actually had the whole span here from screening through referral through care and actual improvement, we would be a lot less picky about the steps.

And, I think part of this is the artificially dissecting this out and for people like myself who weren't here for all the prior discussions Jason brought up, you know, it's like, wow, wrapping your head around it in isolation is probably not fair when I think we have recommendations that this is a worthwhile intervention. It's important.

And, it's more of a thought for NQF than it is for the developer because you're stuck with whatever processes we have.

But, thinking about having a more integrated discussion would probably be helpful, at least to people like me.

DR. NISHIMI: Well, and they could bring back a measure pair the next time. That's always an option. If the other one had been ready, then they could have paired it with this measure and that probably would have been much more helpful.

MEMBER TEUTSCH: I guess, but I heard it's not even their measure, it's a measure from a different --

DR. NISHIMI: No, the one they're developing.

DR. WOODS: There is a measure that exists for the validated screening tool being used. And, then, there's one that we have developed but the challenge with this particular aspect of care, the performance is so poor that when we started with the use of a validated tool, we got -- starting with hundreds of patients, we got down to just like, four who actually got a

referral and got their referral tracked.

So, I mean, so, we were surprised by those results. That's why we didn't bring them forward but we have further help from AHRQ to do further testing. And, we'll do further testing on this and maybe NQF can provide us with guidance on how measure pairs work and how that gets reviewed in a committee or how we could prepare that for you.

MEMBER TEUTSCH: But, you know, it's really helpful to people like me who are not as deeply immersed as -- if we had that whole sweep of what's going on and we really had a good understanding and that's where Arjun was sort of going.

Where is the breakdown in this process? And, I'm hearing from you there are multiple breakdowns.

DR. WOODS: Yes.

MEMBER TEUTSCH: But, where are those breakdowns and then what are the critical measures that we have to overcome?

It would be helpful to figure out, you know, what the -- what to do because, I mean, I know, and I assume I speak for most people here just to -- it's important to get these kids taken care of and because it's important from their perspective, the family, society's perspective.

So, I hate to see sort of rules break down the care process.

DR. NISHIMI: Arjun? Matt? And, then, --

MEMBER VENKATESH: The only thing I was going to add, and I agree with kind of everything that was said there, is that when your next submission in your survey data around face validity testing, it said only 45 percent of people thought that measure was feasible.

In some ways, that's -- I don't know how to actually benchmark that because the vast majority of people don't survey and get data on feasibility. But, I would imagine had we gotten to that part of the discussion, people would have been concerned that less than half of people

actually thought the measure was feasible.

And, I don't know if that's because it's chart extracted or what it is about it, but I would try to get some more information for that part of the application.

DR. WOODS: That's misleading in that we specified both it's a chart review measure and was an eMeasure. And, people did not feel it was feasible and when we tested it, it was not feasible as an eMeasure because the elements necessarily for the documentation of this care isn't -- are not in structured variable fields.

So, I'm sorry for the misleading.

DR. NISHIMI: Matt?

MEMBER STIEFEL: As opposed to a measure pair, I wonder if we could consider a measure bundle of screening, referral and follow up as a measure?

DR. NISHIMI: Yes, developers -- that would be a composite and the developer is free to submit that kind of thing.

MS. MUNTHALI: I just wanted to get

back to Steve's point. I had a conversation with Steve and Matt and I -- we are hearing you about understanding the portfolio and understanding where these measures might fit within a spectrum of health and well-being and population health.

And, so, we hope to have time, if not today, definitely we're going to dedicate one of our webinars to do that so you can understand where there may be gaps in health and well-being and where this measure and other like measures may fit in.

DR. NISHIMI: Okay. Are we ready to move on then to the next measure?

Thank you, Donna.

MS. MUNTHALI: So the next measure under review is Measure 279, bacterial pneumonia admission rate. This is a PQI 11 measure that is stewarded and developed by AHRQ. It's a maintenance measure and just wanted to give you a little bit of background. This measure was initially reviewed by our Pulmonary and Critical Care Committee, and we asked a couple of you on

this committee to provide input from a health and well-being perspective. And it wasn't initially recommended for endorsement by the Pulmonary and Critical Care Committee.

At the Consensus Standards Approval meeting where they review the measures and make sure that we are upholding the consensus standards process, the developers, AHRQ, asked for a reconsideration, and the Consensus Standards Approval Committee co-chairs referred it to the Health and Well-Being Committee. It is a population-level measure.

I can just share with you some of the concerns that the Pulmonary and Critical Care Committee raised. They felt that there was limited risk adjustment on both age and gender, and they were also concerned about no risk adjustment with regards to poverty level.

So the initial votes at the in-person meeting, the measure did not reach consensus on performance gap, validity, and overall suitability. And after it went to comment and

the committee re-voted, they did not recommend the measure. So overall suitability, they voted no.

So a couple of changes since this measure was reviewed by the Pulmonary and Critical Care Committee. The developers have since changed the name to community-acquired pneumonia admission rate, and so we're bringing this measure in front of you.

I just wanted to also let you know that one of the co-chairs of the Pulmonary and Critical Care Committee, Dale Bratzler, is on the phone. And I think he'll be here until about 15 minutes. And Robyn was one of the senior directors on the Pulmonary and Critical Care Committee, as well as our other colleague Reva Winkler.

And I also wanted to note one recusal on the committee. Arjun was part of the developer team, and so he will not be participating in discussion or vote on this measure.

And so, Robyn, I'm not sure if you wanted to add anything.

DR. NISHIMI: No. Dale, are you on the line?

DR. BRATZLER: Yes, I am.

DR. NISHIMI: Great. Was there anything you wanted to say before the Committee began its discussion?

DR. BRATZLER: Well, I'm happy to answer questions. I think there were a variety of reasons that our committee did not recommend it for endorsement, and I'm not sure they've been addressed yet.

MS. MUNTHALI: Okay, great. Thank you. So right now I'll turn it over to Carol Stocks -- hi, Carol -- who's representing AHRQ, and I think you probably have some other colleagues on the phone. So we're asking everyone to give a two- to three-minute intro of the measure, and then we'll turn it over to our lead discussants on the Health and Well-Being Committee.

MS. STOCKS: Okay. I believe we have Sheryl Davies on the phone from Stanford. Stanford is a primary contractor that does a lot of the heavy lifting on the development of multiple -- we have about almost 40 indicators through NQF endorsement.

This measure, as with several others, we call our prevention quality indicators, and they utilize hospital-administrated billing data not to measure quality in the hospital but to measure aspects of what's going on with the population outside of the hospital because, as you know, the cases that come through hospitals reflect a good part of what's going on in the community.

So I think that, because it's not a direct measure of quality of physician care, there is frequently some confusion about what we're looking at. It's based on the concept that, with adequate healthcare resources in the community, a portion of pneumonia cases, community-acquired pneumonia or hospitalization

can be prevented. And it's not a measure of whether appropriate decisions are being made about hospitalization but whether the need for hospitalization occurs.

So of course the concept of access to care is very important. And in the past, access to primary care has been pretty much the sole focus. I think we're expanding on our understanding and there's some research to support this that there may be a lot of other factors going on, perhaps access to appropriate home healthcare services or mental health and substance abuse treatment services, a number of things. And it varies by community, so it's designed to look at communities that have relatively high rates compared to other communities. That could be at the county level, the city, or the state level. It's used by many, many state organizations, public health departments, various organizations that monitor the cost and quality of healthcare in their state.

The risk adjustment is based on, is only including gender and age at this point. We believe that, of course, anyone who uses the software that comes along with the specifications of this indicator can change that, as they need, for their particular use. But we don't have a more sophisticated risk adjustment because we believe that the general concept is that whether a given population has co-morbidities that are more or less prevalent, the healthcare system should be able to rise to the need and meet those needs. So if you kind of adjust away some of that, it's difficult to get a handle on.

And that's for some uses, and, of course, those who are going to compare or do research and want to look at the effects of certain aspects of the healthcare system or interventions that have taken place, then they can utilize more risk adjustment to their needs. I guess that's all I want to say ---

DR. NISHIMI: Great.

MS. STOCKS: --- if that's okay. I've

taken up enough time.

DR. NISHIMI: So Committee discussion, and Emilio and Matt. Emilio then Matt.

MEMBER CARRILLO: I'd like to ask the co-chair of the Committee that's on the phone what issues were not addressed.

DR. BRATZLER: Yes, so this is Dale. So I don't disagree with anything that you just heard from AHRQ, and I'm certainly not opposed in any way to a population-level measure on hospitalization rates for pneumonia.

A couple of things that are our committee discussed. So the overall admission rate for pneumonia has been declining, and the evidence for disparities, this was largely based on variations in county admission rates, but even the developer acknowledged that a substantial amount of the disparities, the actual admission rates, between counties was based on the income level of the population. So it seems --- I kind of get this concept of not adjusting for poverty levels, but, yet, that's what's driving most of

the disparities in the admission rate, even by AHRQ's own submission, admission with respect to the information.

The other thing is I think our committee talked a lot about, you know, even though this is a population health measure and it's all about changes to public policy, community-based interventions that might reduce hospitalization rates for pneumonia, in reality there was no evidence presented in any of the references or anything else that you could actually do anything about those particular policies or any evidence that changing those policies would actually change admission rates. When you look at the reference list that was provided, there were about 23 references. Thirteen of them focus on whether or not to give influenza pneumococcal vaccine, and our committee completely agrees that that's very important. That definitely reduces admission rates, and we have nice process of care measures for multiple settings of care now around influenza

pneumococcal vaccination rate.

But all the rest of the references that were provided were simple observational studies simply highlighting what the differences are that might be associated with higher rates of admission for pneumonia. So certain patient-level characteristics, population-based income rates clearly significantly associated with hospitalization rates. You know, so I guess if it's just a measure to look at population-level measures and compare counties, I mean, frankly, you could highlight that most of the variation between counties is being driven by income levels. And those, as you know, are very difficult to change, particularly for the healthcare system.

And then, finally, I'll make this point. I know it may not be relevant, but it came up frequently in our conversation that there is one unintended consequence of this metric, and it is highly used for accountability at the practice level. And I know that wasn't AHRQ's

intent, but that's what's happening. It's now being used for the value modifier, QRURs, other reports that are being used at the practice level by a variety of payers, not only CMS. And that I know wasn't the intent of the population health measure, but that was certainly discussed extensively by our committee.

DR. NISHIMI: So that would be a usability and use issue for you to keep in mind. Emilio, anything else?

MEMBER CARRILLO: No, I think, in terms of, just to accentuate what you said, the issue of access is central to the rationale for the measure. And income has all to do with access.

DR. NISHIMI: Matt, as one of the lead discussants, and then I'll go to Amir.

MEMBER STIEFEL: Are we going systematically through, are we starting with --

DR. NISHIMI: So we're starting with evidence, if you have any. But, first, if you had any questions for Dale -- I'm sorry --

because he's leaving.

MEMBER HILL: I was just wondering if our colleague on the phone could tell us what the role of antibiotic resistance is, in their view, with this particular measure.

DR. BRATZLER: Yes, so this is Dale again. I don't know that we ever discussed this and, honestly, I don't know, I'm not aware of any specific information about antibiotic resistance driving actual hospitalization rates. One other problem I didn't mention with the metric is that we know, I've been studying pneumonia since the late 90s, and we knew that more than half of patients admitted to the hospital did not have a bacterial culture when they were admitted with pneumonia. And I think what's coming out now through a variety of different forms of research is that there are a whole host of viruses that are causing pneumonia, but I suspect they often get coded as organism unknown, and they may end up being drawn into this particular denominator for this measure.

So, as an example, we know respiratory syncytial virus actually is emerging as a fairly substantial risk for elderly for hospitalization for pneumonia. I don't know of any way to prevent it, and, yet, I suspect that, unless you're doing PCR testing for RSV, the diagnosis is often missed, so these cases get, you know, incorrectly coded as potentially organism unknown cases.

So the epidemiology of pneumonia is simply on the basis that we have better testing now for viral forms of pneumonia is changing. But I'm not aware of any antibiotic resistance issues that impact hospitalization rates. We didn't discuss that.

DR. NISHIMI: Any other Committee questions for Dale? He does have to leave, so I want to make sure you have the opportunity. Matt?

MEMBER STIEFEL: To summarize what I think we've heard is the concerns relate to the appropriateness of adjustment for income; second,

the lack of evidence of the relationship between policy changes and reductions in admissions; and the third is the level of accountability for this measure and the appropriateness of accountability at the practice level.

DR. NISHIMI: Anything else?

DR. BRATZLER: Yes, I think those were the key issues. Again, I mean, I think promoting vaccination, which there's tons of evidence for, everybody agreed with, and there was extensive discussion about literature in the applications. But, you know, the lack of adjustment for poverty was a big concern.

And the other one that I'll just mention, I'm reading back through our actual notes, was acute illness burden that's not uniform across geographic areas. Again, there's no risk adjustment for patient severity of illness or underlying illness. So we know that diabetics, COPD patients, and others have higher rates of admission for pneumonia. What we don't know is, I mean, what's not included in the

measure is that there's no discussion of variations in those actual underlying risk factors based on county-level data.

DR. NISHIMI: Okay. Thanks so much, Dale. Amir?

CO-CHAIR QASEEM: So just a quick question. I think everything has already been addressed about this measure, and apologies for my being late. One question that I had was, I mean, I think we all understand that reducing hospitalizations is important. What I did not see in this measure is that reducing hospitalizations lead to better clinical outcomes for patients with bacterial pneumonia. Is that an assumption that you had behind this measure? I'm not aware or have seen evidence for that.

DR. NISHIMI: Dale?

DR. BRATZLER: Yes. So, yes, that's a good --- I'm trying to drill through my mind here and see if I can think of anything. I don't know that that's, I don't know that there's any evidence of reducing hospitalizations changes

outcomes. I mean, there are a lot of patients that probably -- I completely agree with AHRQ on the point that there are a lot of patients that probably get admitted to the hospital that can be treated in the ambulatory setting using appropriate risk stratification tools. So I certainly agree with that.

And the other thing that AHRQ demonstrated very nicely in their application was that there is substantial variation in hospitalization rates across counties. I mean, there's a fairly wide spread of the admission rates. But I think our primary concern was, was that a big driver of that disparity or those differences between county rates was income level in the county. And from a policy standpoint, I'd love to fix that. I'm not exactly sure how.

But in terms of changing patient outcomes, you know, I think anytime we can keep people out of the hospital is probably a good thing to do. But whether that's been studied explicitly for pneumonia about whether you can

reduce mortality by keeping out of the hospital, I'm not aware of any studies.

CO-CHAIR QASEEM: And that's why it needs an outcome measure, Dale. And my concern was are we going to end up with some unintended consequences, patients who should be getting treated actually now is going to have worse outcomes because now you're trying to reduce hospitalization rates. For an outcome measure, I think that was an important one to have it in there, and that's why I was bringing it up. Maybe it has already been discussed before earlier, I don't know. But anyways . . .

DR. BRATZLER: Well, we certainly didn't discuss that.

DR. NISHIMI: Steve and then Tom. Sorry.

MEMBER TEUTSCH: So as I heard this, this is a community-level measure, not a hospital-specific or clinical-specific measure. And I guess it troubles me a little bit to think that we can't do anything about many of the

things that are related to those income gaps, and I wouldn't adjust for them because they can be addressed.

I'll just give you a simple example. Medicaid expansion. Maybe a lot of these folks don't have access to adequate care because they can't get coverage. Those are policy decisions at a community level. Availability of services. Now, whether the community-acquired pneumonia admissions is the best measure of those things, we could discuss how best to get at them. But I'm not really disturbed about having these things when we need to keep it in front of the healthcare system and the community that there are solvable social and economic approaches, as well as clinical approaches. And we heard about some of them in terms of immunization. We've got a variety of approaches to that that really should be addressed and probably need to be addressed not just within the clinical care system but at the community level because many of these events, I suspect, occur among those people

who are really not within the healthcare system very well.

So I guess, you know, I don't know what the evidence for it is, but I think we have to think a bit more broadly about what we want these measures to do, who is the organization, what are the organizations or entities that need to take responsibility because my definition, at least, of the health system, not the healthcare system, embraces a much broader group of stakeholders, many of whom would not see themselves as healthcare oriented, who need to embrace these things and realize that we're paying the price in many of these metrics, and I'm sure CAP isn't the only one that we would think about at a population level.

And I'd even dispute diabetes and other underlying diseases which have, you know, can be, in large measure, prevented with better clinical care and with more attention to obesity, nutrition, parks, soda taxes, you know, all this stuff. So I think it's important to keep it in

front of us, even though I wouldn't say it's up to the pulmonologist to solve it.

CO-CHAIR McINERNY: Well said, Steve. I have two comments about the hospitalization. One, it increases expense, and so that's a problem and you'd want to try and avoid hospitalization for that reason. And, two, it increases the risk of the patient for having some morbidity and/or mortality from hospital-acquired infections and other untoward events that occur in the hospital. So another reason why we'd probably want to reduce hospitalization for these patients.

DR. NISHIMI: Okay. I think we're ready to start marching through. Thanks so much, Dale. We appreciate your time this morning.

DR. BRATZLER: Okay. Thank you.

DR. NISHIMI: Elisa reminded me that, before we go on, Amir, our co-chair, has arrived. And, Amir, for the record, we need you to introduce yourself and whether you have any conflicts.

CO-CHAIR QASEEM: Sure. Again, apologies from my end. I had a scheduling conflict. Amir Qaseem. I'm Vice President of Clinical Policy at the American College of Physicians. I don't have a conflict, but, I think probably more for disclosure, I am on the board of trustees or regents of directors, whatever their governing board is at the PCPI. But they don't have any measures, so it's not a conflict.

DR. NISHIMI: Okay. So the first criterion that we need to address is evidence. This is an outcome measure, so it is a yes/no vote. This is a maintenance measure, so the previous committee said it was yes and the Pulmonary Committee, in its deliberations, didn't further discuss and vote, so they said yes. And so the question is, is this committee comfortable with just suspending the vote, we sort of had a discussion already, and moving on to the next criterion?

MEMBER SPANGLER: Well, I have a

question. They said yes. I thought they didn't reach consensus and then, when they discussed it, they didn't recommend it.

DR. NISHIMI: They didn't recommend the measure as a whole, but they didn't discuss evidence.

MEMBER SPANGLER: Oh, okay, got it, got it, got it.

DR. NISHIMI: Yes, they didn't discuss this criterion. So then let's move to gap. Emilio, Matt?

MEMBER CARRILLO: There is a gap that has been noted, and, again, the questions, in terms of the stratification of the measure are very important, including socioeconomic status. So we have data provided. County to county, they're significant. That's been pointed out already. There is significant gaps that have been noted and well documented.

MEMBER STIEFEL: And that there are interventions that would reduce this admission rate.

DR. NISHIMI: Any other questions or comments on gap? Are we ready to vote on gap then?

MS. CRAWFORD: Voting is open on Measure 0279, performance gap. One is high, two is moderate, three is low, four insufficient, and we're waiting for 16 votes. Fifteen votes. One more. Everyone hit their clickers again. Oh, then 14 is what we're looking for. Okay, one more. Okay. Can everyone, like, point towards -- oh, that's recused. I'm sorry. Thank you. Well, 14 it is.

Okay. Five voted high, nine moderate, zero low, zero insufficient. Thirty-six percent high, sixty-four percent moderate.

DR. NISHIMI: Okay. So let's move on to scientific acceptability. Matt, Emilio, Amir? Any comments on the reliability?

MEMBER STIEFEL: Reliability testing was done with a high signal-to-noise ratio, 0.97.

MEMBER CARRILLO: I agree.

DR. NISHIMI: Any other questions on

reliability and other Committee comments? Okay. We'll vote on reliability.

MS. CRAWFORD: Voting is open on Measure 0279 on reliability. One is high, two moderate, three low, four insufficient. We're waiting on 14 votes.

Voting has ended. Seven high, seven moderate, zero low, zero insufficient.

DR. NISHIMI: So consensus is not reached, so we'll continue -- oh, it is? Oh, I'm sorry. It's high. I don't have my glasses on. It's high and moderate. Okay. So 100-percent consensus on reliability, so the next question is validity. This is where the Committee might want to discuss the risk adjustment issues that Dale Bratzler raised. Matt and Emilio?

MEMBER CARRILLO: Well, as has been pointed out, the issue of access, which is very important to this measure, is clearly strongly associated with socioeconomic, and there's also the issues of diabetes and COPD that are very critical to the measure that should be included

and stratified. So I think that there are some challenges to validity.

MEMBER STIEFEL: And my opinion, in terms of the concern expressed by the previous committee of the lack of adjustment for SES or income, I disagree with, I think it would not be appropriate to adjust for income for some of the reasons that Steve articulated.

DR. NISHIMI: Marcel?

MEMBER SALIVE: Also, I think, you know, there are some points that weren't made. One is that not just, you know, are there interventions, but I think that highlighting differences in different states, you can look at some of the policies that are in place and whether they would have an effect because, as Steve said, you know, there's different Medicaid eligibilities in different states. So there are some natural experiments that could be looked at with the data, and so I think it's very valuable.

And, finally, I think, you know, at a population level, this type of adjustment is

just, you know, gilding the lily and really is not necessary.

MEMBER TEUTSCH: I would say it's not only gilding the lily, it masks real differences that need to be addressed.

DR. NISHIMI: Anything else? Okay. We're ready to vote on validity for 0279.

MS. CRAWFORD: One is moderate, two low, three insufficient. Voting is closed. Nine moderate, five low, zero insufficient. Sixty-four percent moderate, thirty-six percent low.

DR. NISHIMI: Okay. We'll move on to feasibility.

MEMBER CARRILLO: Feasibility I think has been established by the previous committee, and I don't think that there's any issues there.

DR. NISHIMI: Okay. So Matt agrees. Any other comments or questions on feasibility?

MS. CRAWFORD: Voting is open on Measure 0279. One is high, two is moderate, three is low, four insufficient. One more vote. Eleven high, two moderate, one low, zero

insufficient. Seventy-nine percent high, fourteen percent moderate, seven percent low, zero percent insufficient.

DR. NISHIMI: Okay. And then the last criterion is usability and use. Again, this measure is specified and AHRQ puts it before you as a county-level measure. You heard from Dr. Bratzler that there was, in that committee's view, an unintended consequence of the use of that measure at the provider level. Any other comments, Emilio or Matt or Amir?

MEMBER CARRILLO: I just want to perhaps raise the question is this an issue that's found with other PQI measures, as well, that this sort of slippery slope to a provider focus? If anybody could comment on that.

MS. STOCKS: Could I say something about that? In terms of the CMS use of this measure, and I think that's kind of at the root of the comments, they don't use this measure as we specify it. They have adapted it in two or three different ways with a different denominator

and different numerators. So even though they're called PQIs and they rely on some of our scientific evidence, we consider them a little bit different measure.

CO-CHAIR QASEEM: So that was my major concern when I was talking about the outcome measure. What just Carol said, those of you who are involved with MACRA measures or any of the measures that are not getting implemented, they are approving these measures that say county level but they are getting implemented at individual physician level or level where they have never been tested or we don't have any data that they will improve the outcomes. It goes back to what Steve was talking about. At county level, I absolutely understand for this measure, and you're going to hear me say it many times today because I just came back from a MACRA meeting, as well. It is just very concerning for me how CMS is not only not looking at what level these measures are getting improved, they're not even including the measures that are NQF

endorsed. That's a separate debate, and we're not going to get into it.

I don't have a solution. I want to look at Elisa and the NQF staff. I have been saying this forever, but my worry is that no one is really hearing us out.

MEMBER TEUTSCH: But is this unique to this measure with the QRUR or other --

CO-CHAIR QASEEM: It's a measure-level problem. It doesn't matter which measure it is.

DR. NISHIMI: Steve?

MEMBER TEUTSCH: A couple of thoughts on that because I agree that people misuse statistics, data, all the time, and somehow we can't protect people from that. That's too bad. We can educate them, all that sort of thing.

A couple of things to think about, though. One is the criterion we saw earlier was about healthcare. This should be about a health system more broadly, which might take it at least out of, at least partially out of the healthcare system, certainly part of that. That's one.

The second is this is just one of a portfolio of measures that reflect the general kinds of underlying socioeconomic, health system, financing issues that we can monitor within the healthcare system. And one thing to think about is to put forth a portfolio of these things. So rather than sort of reporting this in isolation, if we had a set of these things that we think are driven largely outside of the clinical care system or at least substantially and actually said, whether you look at CAP or Y or Z, that you see the same phenomenon because the interventions for them are, you know, in many cases, broader. I mean, you could speak about diabetes and COPD, but you know, smoking and diet, those affect lots of stuff and lots of outcomes. Same thing with the payment system and whatnot.

So I wonder if we can, at some point, put out some sort of a portfolio of these that would be at, say, a county, a state, or regional level, whatever it is, that then paint a picture that allows people to get a better idea; and,

hopefully, some of the adverse consequences Amir is talking about could be at least ameliorated a little. But you can't help people from using things for which they aren't intended, but at least you can make steps in that direction.

MS. MUNTHALI: Yes, I think that's a very good idea, actually, as we talk about our portfolio and how these measures may fit into it and where these measures, where the locus of accountability. I mean, this is part of what has happened with how these measures have been adapted for use. And you're very right. It's very much out of our control. What we have asked you to do is evaluate these measures against our criteria on the merits, the scientific merits of the measures, but they're going to be used. And part of that conversation is part of NQF's work around the Measures Applications Partnership. These conversations happen frequently that the measures be placed in programs in which they have been specified for.

With that said, I can tell you we're

having some pretty significant conversations with CMS about, you know, measure use and making sure that the appropriate measures are placed in programs. But it's not just in federal programs. These measures are being adapted at local levels, at state levels for a number of reasons. We're undertaking some work around measure variation, talking, we've started to identify what are the reasons for a variation, why does it happen, how can we mitigate it?

And so we recognize that that's a significant issue in measurement. Not much we can do, but we are starting to actually make some inroads towards coming up with a solution, but it's going to take all of us to do that. Matt?

MEMBER STIEFEL: The issue is the issue Amir raised about using hospitalization as a surrogate measure for the health and well-being measure of the progression of pneumonia. And it's hard to hold the message hostage to its misuse downstream. And so this is a challenging question. It may be the best surrogate we have

for a population health measure of progression of pneumonia, and I think, I guess in my opinion, the misuse issue is an issue that Elisa was talking about in terms of sort of downstream of the measure approval.

CO-CHAIR QASEEM: Just one general comment, Elisa, to respond. Is that a possibility, because some of the richness of the conversations that we have in this committee and other NQF committees, I think it tends to disappear by the time CMS hears about it. Things as of nature that this measure and other measures, too, if we're approving a measure at a community level, there is also discussion happening that this is not being approved at the individual physician level. Shouldn't we have something along the lines that we can have that as an option or something, so we can at least go back to CMS saying -- because CMS says something else. They say, well, your committee actually approved this measure.

MS. MUNTHALI: So they're listening,

trust me. We have our CMS contract leads are listening to this conversation right now. AHRQ is a federal partner of CMS's. I think there need to be some discussions that happen between AHRQ and other agencies that are developing measures with CMS. We're all in it together. We understand, you know, they're trying to put together programs for different settings.

So I can tell you they're listening. We will include this conversation in the report. We're going to have conversations. I asked our contract lead at CMS to join this conversation, so I know she's on it, and I'll be following up with her, as well, because we do know that this would be a significant issue.

But it's not just PQIs, it's not just, you know, this committee. It happens across committees, and it's a big concern of all of ours.

CO-CHAIR QASEEM: So I don't want to derail the conversation. I want to get going onto voting, but would it be okay if I put the

CMS person on the spot and asked them to respond to this issue? I don't know who is on the call.

MS. MUNTHALI: Maybe not put her on the spot. Sophia Chen, are you on the call? Operator, is Sophia Chen on the call? And if she is, could you open her line?

OPERATOR: She has not joined.

MS. MUNTHALI: Okay. So somebody is on here that's listening. But trust us, we'll communicate this.

CO-CHAIR McINERNY: So to follow-up on what Steve I think so importantly pointed out, Rochester has made the claim that they will be the healthiest community in the country in 2020. But do we have any measures that would be able to refute or approve that claim? That's what you're talking about. We need a portfolio.

MEMBER TEUTSCH: So I don't know if you want to get into all this. I can spend a long time on this. So we can discuss what the healthiest community is.

So I was involved with the IOM report

on public health strategies, and we set a goal, and made a recommendation. The secretary set a goal that we would be average in terms of the other developed nations, in terms of cost and life expectancy. I don't remember if we did others, you know, because there's a whole series of metrics. I think it was just life expectancy.

That was an enormous stretch for us to become average by 2030. And, you know, we could discuss what it is, and Matt has a well-being measure, we have life expectancy, infant mortality, maternal mortality. We have lots of metrics we can use. We can come up with composites. I know there are states that are using America's Health Rankings, I know that there are places using county health rankings.

So my concern isn't that we don't have metrics. I think we can find them. I think what we lack is a common purpose and will and a willingness to actually focus on that and what are the major drivers of health, which is about 20 percent in the healthcare system, it's about

40 percent social factors, 30 percent behavioral, and 10 percent environmental. And I think it's unfair to ask the healthcare system to solve all these problems, but if we don't engage the healthcare system in thinking more broadly and helping support those initiatives that deal with those underlying drivers, we're not going to get there.

So, sorry, that's a soliloquy, a little political, but that's sort of where I am in all of this.

DR. NISHIMI: Okay. Is the Committee ready to vote on usability and use?

MS. CRAWFORD: Voting is open. One is high, two moderate, three low, four insufficient information. Three high, eight moderate, three low. Twenty-one percent high, fifty-seven percent moderate, twenty-one percent low.

DR. NISHIMI: Okay. Final vote on overall suitability for endorsement. Is there any discussion? Okay. Voting on 0279, overall suitability for endorsement. One yes, two no.

MS. CRAWFORD: Just one more vote. Twelve yes, two no. Eight-six percent yes, fourteen percent no.

DR. NISHIMI: So the Committee recommends 0279 for endorsement. Thank you, Carol. You had to do this before two committees.

MS. STOCKS: Thank you. Thank you for the lively discussion, too.

DR. NISHIMI: And I think if you, you know, obviously, you heard the Committee had some similar concerns about working with CMS on the communication.

Okay. Next measure, 3067. Patricia, Steve, and Michael had an emergency, and so he's not here.

CO-CHAIR QASEEM: So any comments, Steve or Patricia?

MS. MUNTHALI: We'll actually start off with an introduction by the developer.

CO-CHAIR QASEEM: Oh, sure.

MS. MUNTHALI: And this is CDC. It's a new measure. It's HIV infection screening.

And, Abigail, if you can turn on the mike.

MS. VIALL: All right, okay. I can figure this out. Hi. So I am Abigail Viall. I am from CDC's National Center for HIV/AIDS, Viral Hep Atitis, STD and TB Prevention. You saw one of my colleagues yesterday, John Ward. And we are here to discuss, actually our center has put forward both the HIV screening measure and the measure that follows, the viral load suppression measure.

The HIV screening measure is an eMeasure. It is intended to improve implementation of CDC's and subsequently USPSTF recommendations that all persons between the ages of, for the USPSTF, 15 and 65 should be screened at least once for HIV in their lifetime. It is an ever measure, and that has proven challenging in discussions in the past. We have had a great deal of deliberation before we ever brought this forward, and I'm eager to engage with the panel on that. But we feel that, given the fact that, despite having had a CDC recommendation to this

effect for over ten years and the USPSTF recommendation similar to this for at least three years and the fact that the rates for people who have ever been tested for HIV continue to just inch up slowly, that having a measure that would help people with the implementation and help sort of focus implementation of this recommended service would be helpful.

And I believe I have several colleagues on the phone. Will they be able to get through? Okay. It's Cal Ham, Eileen Wong, Wendy Lyon, Kaijie. I think that, for this measure, that's enough.

DR. NISHIMI: So, operator ---

(Simultaneous speaking.)

DR. NISHIMI: --- could you make sure those lines are open?

OPERATOR: All lines are open.

DR. NISHIMI: Thank you.

MS. MUNTHALI: So Steve and Patricia.

MEMBER TEUTSCH: I think the evidence that, you know, testing and getting people into

care, is obviously there, we have those recommendations. What I'm just wondering about was, I mean, we'll get to the ability to measure this in a minute, but to what extent will we have the same question about referrals as we had before? Just testing obviously isn't adequate, right? You've got to get people into a process of care. You were sitting here, so you heard the discussion and I'm not going to repeat it, but it's largely the same set of issues: how do we know that then, after testing, to what extent do the right things actually happen?

MS. VIALL: That is a good question, and I did hear the discussion. I can say that, from our surveillance data, we currently believe that at least 70, I think 73 to 74 percent of people are linked to care, linked, not referred, linked to care within three months of their diagnosis. This has been a huge push in the HIV community to make sure that diagnoses lead to linkage, and, in fact, most states are now using their surveillance systems to check if a

diagnosis, when they receive a diagnosis, it is followed by a viral load or CD4 measure within, well, now, actually under the new NHAS update, within one month after diagnosis.

So we're monitoring that at a public health level. At the measure level, I will be honest. When we developed this, we had a lot of back and forth about what goes in and what goes out and how that affects the feasibility and likelihood of implementation, and documenting referrals is still very tricky, especially in EHRs. And so we could develop a subsequent measure. We actually have other measures that look at retention in care, but we've also seen that they don't get the uptake.

And so it is a thorny issue that gets to the heart of where EHRs are today, and so our emphasis was to get the testing done because in other arenas we have focused on the linkage. Again, referral for us is not enough. When we talk about what happens after care, it's linkage, and linkage for us means they have seen an HIV

care specialist. So we have been trying to tackle that through other options.

MEMBER TEUTSCH: I'm hearing there's at least 30 percent falling through the cracks because even that 70 percent, they don't necessarily follow through either.

MS. VIALL: If you look up to a year out, it's actually much less than 30 percent. Again, our emphasis is increasingly not on just linkage but fast linkage. Linkage, you know, within three months used to be the barometer. Now we're moving to one month. And so the goal under NHAS, I believe, is like to get 70 percent within one month. So it's not just linkage, it's immediate linkage. It's an emphasis on quick linkage.

So I believe if I looked at, like, the 12-month numbers -- if any of my colleagues are on and have that data in front of them, they're free to chime in here.

DR. NISHIMI: If we could just return to the evidence.

MEMBER MCKANE: That's kind of where I was going because, I mean, this is all a great discussion, but the measure, I mean, first you have to have screening. I mean, this is like a setup for other measures to come. And, you know, there are, the task force recommendations are for screening of adolescents and adults aged 15 to 65, so I think that they've, from what I can read, they've provided, you know, systematic, you know, good evidence that was graded for the importance of screening.

The other facets of this is, I mean, I think that discussion comes into play at some point. But I think that the focus for this measure is on the initial screening and the fact that there are gaps, which is the next discussion, so I'll try not to get ahead.

DR. NISHIMI: Any other comment on evidence, per se?

CO-CHAIR QASEEM: So I just have a couple of comments. First of all, I think it's a very timely measure. I absolutely agree with you

the HIV screening has been, we're not still doing as good of a job as we should, considering the evidence that is out there.

Not a criticism, I feel like this measure actually doesn't go far. I was a little bit concerned about the 65 as the upper age limit, as well, and there is a lot of new emerging evidence that's questioning 65, especially folks who are living in other living situations. So that's one issue.

And then I'm concerned on the lower age limit, the 15 to 18, because that always is a tricky one for, you're not an adult. And my worry is will that be under physician control? Is there any evidence that you have that we can still screen, despite all the issues that are associated with the younger population?

And the last comment that I had in terms of, again it's based on the evidence is that I've always, and I know there is no evidence, Abigail, and you probably won't be able to answer that one anyway, but this one-time

screening just bothers me a lot because a 55-year-old adult man who was screened 20 years ago would satisfy the criteria for this measure.

So it's sort of a general comment, if you have any response. Otherwise, as I said, this is a very timely measure. It's about time. I've been pushing CDC for this for a long time.

MS. VIALL: Well, good. Then I'm glad we're responsive eventually. So I will take those in turn for the over 65. I think there is a lot of interest in re-examining that upper bound at this point because of the way the USPSTF guidelines are developed, the way our guidelines are -- I'm not sure we could have a measure that goes beyond the evidence and get good uptake.

The other issue is, particularly in the over 65, I'm not sure, although a lot of the studies that were done earlier looking at this are old at this point, but the question is whether it would make sense to do general screening or targeted screening, which gets to what CDC ultimately hopes to bring forward in

future years is targeted testing measures based on risk. Those have proven very thorny, as well, because cataloging and documenting risk in electronic health records or administrative data is very difficult. So I would say, at this point, we do have other measures in the works that might address the over 65, at least to some extent, based on risk.

For the 15 to 18-year-old population, that is a challenge. We do know that, I want to say there was recent BRFSS data -- and, again, folks on the phone, if I'm mangling this entirely, please jump in -- but I think it was somewhere along the lines of at least 30 to 40 percent of sexually-active adolescents between the ages of 15 and 18 had at least been tested once for HIV already. So it's not great, but there are ways to get to improving in that particular population, and we would say that that's a group that we really want to emphasize because a lot of the incidence right now is among younger people, not --- adolescents and young

adults, 18 to, say, 24. So that is less a problem with the measure than how CDC tackles implementation working with its partners.

Finally, for the one-time ever, this is the hardest part of this measure, both conceptually and in terms of actual implementation. At some point, I think there is probably room now to go back and look again, but the evidence right now, for one-time only, it's very strong on both an individual level and a sort of population cost-effectiveness evaluation level.

Again, given the changes recently in treatment guidelines, it's possible that a more frequent, more recurrent screening rate for the general population could now be cost effective, but it hasn't been looked at and, again, the way guidelines development and implementation processes go, it could be several years in the making. So we're trying to push the one that has been accepted as valid.

MEMBER HILL: Yes, can I ask if you

can elucidate why you selected the age 15 as the lower end when the CDC recommends 13?

MS. VIALL: So CDC would say 13. To be honest, because the USPSTF recommendations have more sway, they have more sort of established power -- especially now in law -- than the CDC recommendations do. And also because some of the groups that we would want to work with on implementation are a little bit more reticent about endorsing for the 13, 14, and 15 than they are for 15 and above. So groups like American Academy of Pediatrics -- or American Pediatrics Association, folks like those, some have been more reticent about that 13 to 15, whereas they've been more open to adopting 15.

But, yes, CDC -- had it just been in our power and we had the same cache as USPSTF, we would have gone for 13. But, sadly, we don't, at least according to the health reform law.

CO-CHAIR McINERNY: So one of the problems with testing at 15 to 18 is patient confidentiality. Because so many times, although

you and the patient may have confidentiality and the parent is not in the room, etcetera, the insurance claim forms come to the parents and they see HIV testing and say, well, what the heck is that all about? And that can be a big problem for the pediatrician. So I think to get above a certain percentage in that 15 to 18 age range is going to be difficult.

MS. VIALL: And I will be honest. CDC has not only our division of HIV/AIDS prevention, but also our division of STD prevention. This is a big issue for testing for STDs and other certain confidential services that kids don't necessarily want their parents to know they're getting. And so we have been working with the Guttmacher Institute and others to sort of examine the issue of EOBs and whether there are ways to structure the way EOBs are handled for certain services in certain vulnerable populations, which, technically, is not just adolescents. It could be wives of abusive husbands or certain other populations. So

dependents, in general, are potentially vulnerable when they get certain services, and so looking at the EOB issue is a separate policy issue that we are looking into.

MEMBER HARRIS: I'd just like to say that in Tennessee part of the work that I've worked with the Department of Children's Services are with children who are part of the juvenile justice system and they come into this system at 12 and 13 and all of them have HIV testing/screening, period.

Part of the work that I did with the FQHC -- where I was previously chief medical officer -- the way we handled it was that we had it as, you know, a bundled conversation versus it being separated and teased out. And as a part of the initial paperwork for signing in, there was the notation related to opting out versus us having to have a conversation about opting in.

So I'm not sure if there's a way that this could be a part of the measure. But, you know, typically, that was the way that we handled

it.

MS. VIALL: And those are in keeping with CDC's recommendations in 2006 where we did emphasize that this should be routinized, it should be opt-out, it should be part of general consent. And, really, we see this measure as furthering that because it's sort of standardizing this. HIV screening is like cervical cancer screening, it is like breast cancer screening, it is like any generalized screening. And having a measure that kind of says that that is something that we think physicians should do and be held accountable for is part of moving towards that conversation. But I think there are other implementation issues that we've been working a lot with state health departments and providers on to sort of smooth the process for all groups so that people can get these tests.

DR. NISHIMI: Jason?

MEMBER SPANGLER: I may have missed this in the paperwork, but do you have a

definition of screening in your measure? Because it seems like testing and screening are being used interchangeably, and I'm wondering if the measure should be testing, not screening, because you say in your numerator it's either documentation of a test or evidence of HIV infection, but you also say that HIV infection is often not recognized by physicians. So I'm not sure if I think that the evidence of HIV infection should be part of the numerator unless you have a definition of screening, aside from testing.

MS. VIALL: This may be a CDC internal baseball semantics thing. So we have often used screening when we talk about generalized, not dependent on risk, and we use testing internally when we're thinking more about risk-based or diagnostic testing. So I don't think that we would be opposed to making it a testing measure, as opposed to a screening.

The other thing, though, is that the USPSTF recommendation states screening. So this

sends a signal in terms of we are aligned evidentiary-wise with the USPSTF.

MEMBER SPANGLER: So I understand that completely. But then my question would be the second part of that numerator, evidence of HIV infection, how would that be determined? Apart from testing.

MS. VIALL: ICD-9, diagnostic codes. I mean, as part of EHR core elements, your diagnoses -- your current diagnoses should be documented in the record. And then also ICD-9s are in there. Or ICD-10 now, but --

MEMBER SPANGLER: So you would be okay with an ICD-10 code but no evidence of an HIV test?

MS. VIALL: It seems unlikely that a person would be receiving active care for HIV who hasn't received a test. So I think, in that particular case, there would -- having another test for that person -- I mean, this measure already has the potential for over-testing. Let's just put it out there. So in that

particular case, the potential that you're doing unnecessary testing really, really skyrockets.

MEMBER HARRIS: I guess I just wanted to get further clarity. So when you're saying screening, you're talking about the OraSure, you're talking about just a simple blood test to determine whether or not someone is HIV positive. But then when you're talking about testing for HIV, then you're actually speaking about the actual RNA test for the virus itself, or what exactly are you using for your criteria?

MS. VIALL: When we say screening, we mean that the person has received testing according to CDC's 2014 testing algorithm, which is now sort of a standard for the United States. So that would be the initial Ag/Ab test and then an HIV-1/HIV-2 differentiation based on the results of the first screen, and then there's actually a third step, too, depending on how the HIV-1 and the HIV-2 tests go out. So we expect people -- when they're tested or screened, that they're getting the algorithm for the whole test.

MEMBER HARRIS: Was that actually listed in this document?

MS. VIALL: It is not.

MEMBER HARRIS: Okay.

CO-CHAIR QASEEM: So, Abigail, can I follow up on what Jason just brought up? I think it's an important issue. So the task force does talk about the testing part. The second part that you have added on, the evidence of HIV infection, am I hearing correctly that you're willing to delete that part from the numerator? Because I'm not aware of any evidence -- and if you're talking about the evidence, I'm not sure if that leads to improved outcomes, depending on what stage that's happening in. All that gets a little complicated. We're talking about screening, which is through testing.

MS. VIALL: So the reason we added that in the numerator was because the assumption is if you've been diagnosed with HIV and you're being treated for it, you have ipso facto been tested and diagnosed at some point. So this,

again, is getting to the fact that healthcare records are fragmented. And so we wanted -- and, again, it is a one-time only measure. So we included that in there to minimize over-testing.

Another option would have been to just exclude them from the denominator. That's another way to handle people who are already diagnosed. The reason we wanted to include them in the numerator was because, at some point, that person got a test, so they met the measure at some point. And so we didn't want to subtract credit for that group.

But, you know, you could reconstruct it taking them out of the denominator. We just wanted to acknowledge the fact that a person who has been diagnosed has been tested and so, at some point, met the measure.

MEMBER SPANGLER: So you're saying those patients who were tested but might not have been documented that they were tested, but later they had been diagnosed with it somehow, is that -- I'm trying to not let anyone fall through that

--

MS. VIALL: Yes. I mean, so, basically, again -- because, you know, in an ideal world we would all have our longitudinal records and they would be great. But, you know, a person that was tested for HIV and diagnosed in, say, 1990 and then they've been getting care ever since, the record that this measure is being sort of run against the algorithms, they may have the diagnosis and the documentation of care in their current record. They may not have that test from 1990. But we want to give credit for the fact that that person was tested and the doctor knows that they were tested. So, yes, that's why we included it in there.

MEMBER BIALEK: I'd like to follow up on Barry-Lewis's question and the response to the question. It was very helpful having the specifics for the measure noted. From an NQF standpoint, two questions. One is, for all the measures, it would be helpful if we had the specifics, if the developers provide those

specifics. And, secondly, if we endorse the measure, can we ask that those specifics be included? Because the measure itself lack that specificity. It's in the spreadsheet? Okay.

MS. VIALL: Well -- go ahead.

MS. MUNTHALI: So if the specifics are not there, we can ask the developer to update the measure with that information --

MS. VIALL: I'm sorry. I lost -- who was the person and what information did I give?

MS. MUNTHALI: He's talking about, generally, specifics to clarify --

MEMBER BIALEK: What was included in the test? So what --

MS. VIALL: Oh, the testing algorithm? Oh, yes, we can easily --

MEMBER BIALEK: But Marcel said -- you said that it is in there. I thought you said it wasn't in there, which is why I was bringing that up. I'm so confused.

MS. VIALL: So I would say it's not in the measure at this point because CDC's algorithm

is supposed to be the standard for testing. We would say that any test would need to, ipso facto, follow the algorithm and laboratories should know that. But could we put that in the sort of detailed specifications for how this is implemented? Certainly.

MEMBER SALIVE: Well, I was just going to make the point that, you know, every screening measure has this issue that, you know, there's false positives and false negatives. And so they have, I think, done it properly here, so I have no concerns. But, you know, we'll get to that on other screening measures, as well, that you can't -- yes, you kind of have to translate it because you don't want, you know, you want that next step, it's a natural next step of validating the screening test. So I have no concern.

MEMBER TEUTSCH: I want to get back to your point about including the people who already have an HIV diagnosis in the numerator and the denominator. It seems to me that that creates issues with the purpose of the measure because if

you're in a place that has high HIV rates you're going to have -- that's going to sort of boost the -- if I went into an HIV clinic, everybody would have HIV, right? And they'd get 100 percent when, in fact, they have nobody really, other than people coming in, you know, for screening. But if they're really a treatment clinic they'd have everybody positive and, yet, they don't even do -- they wouldn't be involved with this.

So I wonder if there isn't a virtue in actually eliminating it from both the numerator and the denominator so that you're actually looking at -- what we really want to say is, of the people who are out there and don't know they have it, have they been properly tested, in which case you'd get to a different specification.

MS. VIALL: So, again, I'm not diametrically opposed to that in any way. The reason we had the numerator broken out that way is so that you could actually look at it. And you will see, also, in the validation testing

that we provided you, we did run this in -- one of the CHCs that participated in the testing was a Ryan White Clinic, and so their performance rate is substantially above that of their FQHC brethren.

The reason, again, that we did this was because we still consider -- if that provider didn't -- I guess -- let's see, the best way to put this. We still wanted to give people credit for knowing the status of their patients, whether they knew that status because they already knew that this person was coming to them for treatment for HIV or they knew it because they had run the test. Either way, that provider has done the right thing. They have ascertained the status of their patient and know that that person has been tested at least once.

It was mainly a philosophical thing. We had a group that helped us develop this measure, and it was kind of like do we want to "see" the people that have been diagnosed with HIV in this measure or not, and the group

consensus ultimately came in that, yes, we would like to see them in there, but we've constructed it in a modular fashion so that you could differentiate is the high performance because it is a Ryan White Clinic, or is the performance driven by the fact that this is a primary care clinic that is ascertaining the status of most of its patients? So that was why we actually have the three numerators. But, again, it was a philosophical choice.

CO-CHAIR QASEEM: Sorry. Last comment and then we'll move on to the next topic.

MEMBER MCKANE: I agree. I see there's a benefit in knowing, you know, the totality of HIV infection, but I like that you can parse this out so that we can answer the questions that Steve raised.

CO-CHAIR QASEEM: There you go. It's on. The other light was on. Sorry. So we're going to, we will have to vote on this measure eventually, once we're done with the discussion, based on what we have been presented in front of

us. And if we do want the numerator statement to change, then we will have to go back to CDC, and then CDC will have to come back with a modified version. So we'll get to it.

So moving on from evidence to gap? Oh, sorry.

DR. NISHIMI: We need to vote on it.

CO-CHAIR QASEEM: Oh, we just keep on voting on each section?

DR. NISHIMI: Yes.

CO-CHAIR QASEEM: Oh, okay. All right. So in that case, what we have in front of us is what we have in front of us. So we cannot really modify it, so today what we need to do is decide whether we're comfortable with what we're being asked to vote on. So let's vote on the evidence.

MS. CRAWFORD: Voting on Measure 3067, evidence. One for high; two, moderate; three, low; four, insufficient, and we need 15 votes, please. We have 10 high, 5 moderate, 0 low, 0 insufficient. 67 percent high, 33 percent

moderate.

DR. NISHIMI: So I think we can move to vote on gap because we've discussed that.

MS. CRAWFORD: Voting is open for performance gap on measure number 3067. One, high; two, moderate; three, low; four, insufficient. We have 12 high, 3 moderate. 80 percent high, 20 percent moderate.

DR. NISHIMI: Okay. Now we move to the specifications and the reliability, and we've had some discussion about that already, about the numerator. And there was also discussion in the staff evaluation about the denominator, lack of denominator testing. So if the Committee could discuss that, and then we'll be ready to -- and I guess Patricia and Steve.

MEMBER MCKANE: The developers did do reliability testing. It was data element. And with the NQF assessment that followed the specifications, you know, they landed insufficient. And when I was reviewing this, some of the issues that I saw, and also there's

some comments in here that are very good, that they didn't test -- they tested the numerator only and not the denominator, although that may or may not be an issue. And the study was limited to the Chicago area. There's no geographic variation at all. This time, it is the Midwest, so go. And so there might be some concerns about generalizability across the country, but there were different types of health centers, which that's excellent.

But I thought that, you know, I wasn't really clear on the reliability testing, and there were some questions when I was looking at this. It's probably okay, but empirical evidence is very limited. And so I don't know if other committee members have comments, you know, that want to get in on this or not. I do think there's some questions about the reliability testing.

MEMBER TEUTSCH: Yes. So to put a little finer point on that, so there's a testing of the elements in the EHR that I, frankly, don't

understand but apparently pass muster by those who do. But I actually didn't, would have liked to see a little bit more explicit description of things like opting out, how that was handled and whether that's actually in the record. I think you related to the problem of how do you really know if somebody had a test 40 years ago, and it sounds like you don't want self-report; you want to have more documentation, which, in our current healthcare system and our population mobility, is almost impossible. So it struck me that -- and it's really hard to check the reliability of that.

So it struck me that there were some significant reliability issues here, and, you know, I was struck also by the fact that this was sort of in one, in Chicago largely, FQHC-oriented population. I don't know if that's good or bad, but it did strike me as relatively limited. So those were some of the things that concern me.

MS. VIALL: So can I take this? Okay. So let's go through those in turn. The measure,

there's no handling of opting out because opting out is not considered in this measure. It's just like when you talked about cervical cancer screening yesterday. Our feeling is a person may refuse a test at some point for various reasons, but this is recommended, and so physicians should try to understand why a person is refusing a test and then work to address those issues. So we do not make an exclusion exception for opting out, so that's not something that -- it would be very hard to document in the EHR anyway because that's not a standardized national sort of data element. So it would make the measure very hard in terms of feasibility, but it also, philosophically, is not in keeping with our expectations of providers, vis-a-vis screening for this service.

Let's see. I will mention the reliability and validity, it was a little confusing, honestly, for us, too. We had to work with NQF staff multiple times to understand it. So because we did data element testing, they didn't ask us to fill out the reliability

section, so some of these measures, some of these issues may not have been addressed as well in there because we just moved to the validity section.

So that handles opt-out for the one-time only. That is a large issue, and it came into, you know, CDC has vacillated a while on developing this measure because of our concerns about over-testing. When the group that developed this measure sort of met, we did -- and I will be honest, we never documented them, but we sort of did back-of-the-envelope calculations where we looked at what are the costs, individual and for the health system, of late diagnosis of people not knowing their status, and how do those pan out with what we think might be the missing data in the record, in terms of the potential for repeat testing. And what we ultimately came to the conclusion of is that the costs of the repeat testing are not, they're not as great as the costs for the individual in a health system of late diagnosis or missing a diagnosis.

The other thing that we would say is if a provider does not know the status of his or her patient and cannot find documentation, he probably should test again. And there was a lot of philosophical debate about self-report, but there is evidence out there that a lot of people just figure I've had a blood test by my provider in the past, and surely he or she must have already tested me for HIV, so that's why we didn't go with self-report as good enough because of that issue.

So, yes, there is a potential for over-testing here. Again, weighing the costs and benefits, we thought that the value of getting everyone tested outweighed the potential risk of some over-testing.

CO-CHAIR QASEEM: Any other comments on reliability or validity? Because I do want to keep us moving. We're a little behind on our agenda.

MS. VIALL: Can I address the Chicago thing?

CO-CHAIR QASEEM: Okay. Summarize your comments, please.

MS. VIALL: Very quickly, CDC has an interagency agreement with CMS. We are going to do additional testing for this measure. We intended to do additional geographic testing and also in different health systems.

When I brought this measure forward to the NQF staff, I asked should this be a time-limited endorsement or sort of a trial use, like you've done for the Hep C measures yesterday, because we do intend to do additional testing, or is the testing we've already done good enough? And they said the testing you've done already is more than most measures come forward with, so, even though it's geographically restricted, so please bring it forward for full endorsement and then just note that, yes, we have given money to CMS to develop to do additional testing for this particular measure.

CO-CHAIR QASEEM: So Elisa and then Arjun, last comment. We're going to vote then.

MEMBER VENKATESH: Thanks. So I'm not particularly, I guess, as concerned about the opt-in/opt-out. The way I think of this is this: there is insufficient data on reliability. I think that is okay because this is specified as an eMeasure. There's not going to be a lot of available test beds or things to get a lot of reliability testing done, and so we go down this path where we go to then look at the score element validity. The only one that's concerning to me there is this question of are we, can we validly capture previous testing that may not have happened within that EHR, previously done, things along those lines.

I think you guys have an adequate answer as to why you think you understand the risk of over-testing and that you've thought about that, done this back-of-the-envelope calculation, and so it's still valid to do it this way.

My primary concern on this measure is what we call threats to validity section, which

is at the end of this. And I don't know if this is going to fit into this vote or the next vote. But the question of a threat to validity is, is the measure score show meaningful differences in performance, is the question. And the reason that's so important is because you guys are working with CMS on this measure and indicate in use that it's for accountability programs. And so you want to use it to compare two physicians, two facilities, potentially use it in MIPS, as you have in your application. And that means that I have to be able to compare two scores directly the way they would get used.

And so the problem with the measure is this numerator issue. If you do not exclude those with HIV, then somebody's score of 40 percent cannot be compared to somebody else's score of 20 percent for accountability purposes. And that is not a meaningful difference in quality. That is just potentially a meaningful difference in HIV prevalence. And that, to me, is a massive threat to validity.

And so I think, as it is currently specified, if it is going to say that the numerator is either HIV status or that you got screened, then it does not actually reflect meaningful difference in quality between two, whatever your level of measurement is. And that, to me, means low for a threat to validity, which if we, since we don't have reliability testing, we have to vote on validity, I guess I would vote low for that reason.

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

MS. CRAWFORD: Voting is open on Measure 3067, reliability. One is high; two, moderate; three, low; four, insufficient. We have 0 high, 5 moderate, 5 low, 5 insufficient.

MS. MUNTHALI: Well, so we just wanted to let you know. So this measure fails on reliability because less than 41 percent have voted high or moderate. So --

CO-CHAIR QASEEM: So, Abigail, you have probably heard enough comments from the

Committee. I don't think it's going to be too difficult to fix this measure. The Committee is generally supportive of it with some issues with the numerators, denominators, and all that you already heard. So I think it would be very helpful if you could bring it back again for reconsideration.

Okay. Moving right along, Measure 3086. No, 3087, sorry. No, I do have it right. 3086. It's also a CDC measure.

MS. VIALL: Hi again.

CO-CHAIR QASEEM: Do you mind if you can just keep your introduction to as short as possible, please?

MS. VIALL: Yes, yes.

CO-CHAIR QASEEM: Thanks.

MS. VIALL: So, in essence, this measure is intended to look at state performance with respect to achieving viral load suppression among people living with HIV. Viral load suppression is a good barometer of whether we are meeting the both individual needs of people

living with HIV, since viral suppression is a good indicator of whether they are sort of healthy and more likely to live longer lives. It's also a huge indicator of transmission, or ability to transmit I guess is more the case.

So this measure addresses both an important public health aspect from the individual patient perspective and from the public health perspective. It is a complement to an individual or provider, individual provider/clinic-level measure that HRSA already has NQF-endorsed. So getting to our previous discussions about the desirability of having sort of community-level and provider-level measures, this measure is intended to complement and sort of extend the existing measure, or the existing NQF stable around HIV measures.

CO-CHAIR QASEEM: Thank you. Steve, Emilio, any comments? Any general comments from the committee?

MEMBER HILL: One question. This seems to start at age 13 instead of 15.

CO-CHAIR McINERNY: Yes.

MS. VIALL: That's because we own this. This is actually, it's based on our surveillance systems, and most of our surveillance systems, we look at pediatric HIV/AIDS separate from adult.

CO-CHAIR QASEEM: So can I ask just a general question? Maybe I didn't really understand this. This is just pretty much, you're just collecting data, so why can't you just get this state-level data? Why does this need to be a measure? What's the point behind this measure? I'm not getting it. This is statistics, essentially.

MS. VIALL: What do you mean, what's --

CO-CHAIR QASEEM: You're going to get state-level data. Well, you can just collect this data. Why does this need to be a performance measure?

MS. VIALL: So that is, that was actually a question that we all asked ourselves

internally. I mean, honestly, CDC can keep doing this measure. NQF has expressed interest, and, as far as we understand from the IOM, there's interest in starting to consider measures that are not just at the provider level, but the idea of measuring quality sort of across strata, strata of performance, so the individual provider level, the clinic level, the plan level, the community level.

So we're a little unclear on where this measurement area is going, but we thought we had a good measure that sort of fits what is happening in this space, and we put it forward because we understood that the call of this particular committee was community health measures, and we think this is an important one. It is the sort of signature one being tracked by the National HIV/AIDS Strategy. But, yes, we can do this whether you endorse it or not.

CO-CHAIR QASEEM: Yes, because it's a performance measure, so I'm trying to figure out what are we going to be -- who is going to be

improving what with this. As I said, it's a very crude way to say it. I look at it as statistics.

MS. VIALL: States are being tracked on this already, and we are pushing performance increases. This is -- state performance on this particular measure is, again, one of the number one priorities in NHAS. So we are tracking this, and we do expect states to improve, and they have already shown trends towards improvement.

MEMBER VENKATESH: I guess the state-level use I could think of is state Medicaid agencies or states that apply for SIM models to CMS have to propose quality measures for different waivers and for different programs. And so I imagine this could be a measure they would choose as an intermediate outcome or an outcome measure of state-level efforts in those programs.

MEMBER CARRILLO: I agree. I mean, I think that there's value in states competing with each other. I mean, it's a NQF measure. It's not just a, you know, statistic in a book. So I

think that it makes sense for us to move forward with this.

CO-CHAIR QASEEM: So let's go through the process of voting, unless there is any other further discussion on this measure. I'm not seeing anything, so evidence.

MS. CRAWFORD: Sure. Voting is open on Measure 3086 on evidence. One, high; two, moderate; three, low; four, insufficient. All votes are in. 5 high, 10 moderate, 0 low, 0 insufficient. 33 percent high, 67 percent moderate.

CO-CHAIR QASEEM: So we're going to go with gap now. Performance gap.

DR. NISHIMI: Any discussion on gap? I think someone raised the question about why it was only 29 or however many. 27 states.

MS. VIALL: It's now 33 in this year's report.

CO-CHAIR QASEEM: Well, I have a question. Performance gap for who? Who are we measuring here?

MS. VIALL: States.

CO-CHAIR QASEEM: Do state-level performance measures fall under the purview of NQF, something along the lines they do?

MS. VIALL: Yes.

CO-CHAIR QASEEM: I just wanted to make sure. Maybe I missed that.

MEMBER TEUTSCH: So, I mean, there are two gaps, right? I mean, one is the number of states that are doing it, but it's also that there's a big gap in viral suppression.

MS. VIALL: Yes.

MEMBER TEUTSCH: All right. And so there should be some accountabilities for --

MS. VIALL: Yes.

MEMBER TEUTSCH: -- closing both those gaps.

MS. VIALL: That is CDC's position. And, you know, what we would say is, you know, again, in the National HIV/AIDS Strategy, the goal is that it's 80 percent, and no state, in my memory, has yet met that.

MEMBER MCKANE: I have a question. There's only a handful of states that have been doing this. Are there plans to expand this to all the states? And then is there data available at, like, a sub-state level or in small geographies? Because we certainly have lived that experience of needing to know something that may look fine at the state level is not good in certain areas of our state.

MS. VIALL: So I would say it's more than a handful. It's over half, 33 states out of 50. Well, and D.C., so let's call it 51. And, yes, there are, states are moving in this direction. This is a huge push for CDC and for states in general to be able to measure viral load suppression among the residents. But for most states, this begins with passing laws, passing laws to make sure that viral load and CD4 count data, all values, all test values are reported to the state surveillance program. That's the first step.

And states have been rapidly pushing

for changes in their laws. There are very few states at this point who don't have laws in place that mandate reporting. So that's the first push.

And then there's a push for actually setting the processes up and making sure that you're getting all the reports from the labs and making sure the data are high quality. CDC does not report until states have not only gotten the laws in place but then have also done the quality sort of assurance. And so that's why it's 33 states with mature systems. There are many more states that are getting these data and could probably calculate it. We just don't have as much, I don't want to say faith, but we're still a little bit less certain about reporting this. Those states can calculate and use it for their own purposes, but for a national report, we're not quite putting them out yet.

CO-CHAIR QASEEM: Ron, do you have a comment?

MEMBER BIALEK: Yes. You know, I

think this measure, we're talking about states, and I think we're talking far beyond state health departments. We're really talking about the health system, talking about policy. We're talking about really the role of the healthcare providers even in policy development and policy advocacy, et cetera. And I think that's part of what we're supposed to be doing as a committee, dealing with population health measures that are outside of the clinical, that may have a policy component to it.

And so I just want to make sure that we don't get wrapped around the clinical pieces on this because it is broader than that. It does require the policy intervention and the responsibility of the health system providers in a state to engage in that policy development.

CO-CHAIR QASEEM: Any other -- okay. So I'm not seeing any other comments. So let's vote on it, please.

MS. CRAWFORD: Voting on Measure 3086, performance gap. One, high; two, moderate;

three, low; four, insufficient. 10 high, 5 moderate, 0 low, 0 insufficient. 67 percent high, 33 percent moderate.

CO-CHAIR QASEEM: So let's take on scientific acceptability. Can we take reliability and validity together, please? Any comments? Okay. Seeing no comments, let's vote. Oh, there is a comment. Steve?

MEMBER TEUTSCH: Well, I'm worried about a couple of things here. One is the system is actually pretty slow.

MS. VIALL: It's getting faster.

MEMBER TEUTSCH: It's getting faster. Having coordinated surveillance at CDC, I understand the problem. But it is a problem because it lags by two to three years, actually, in some states, which is a long time when you're trying to do quality improvement.

Along those lines, I guess I'd make a suggestion that there be interim reporting because that would then lead people to get their act together quicker and get the data to you and

get it out faster, and so it can be more reliable. More usable, not necessarily more reliable.

CO-CHAIR QASEEM: Matt, you had a comment?

MEMBER STIEFEL: Well, we should be, the staff rated this as insufficient, and I think it's worth a little bit of discussion and a question for the staff about the algorithmic insufficient rating because, otherwise, we may fall into exactly the same issue as the previous measure.

DR. NISHIMI: So they didn't conduct empirical testing at the data element level, which could have been used for the reliability testing. For the reliability testing, they cited the systems and quality control for their data that, you know, the data inputs, if you will. And under the NQF algorithm, that becomes an insufficient rating.

The committee can obviously decide that, you know, based on its experience and

knowledge, the inputs and then follow up into CDC there's a continuous stream or at least mostly continuous stream because there is some places that enter it manually. And so they could rate it otherwise, but that's what led to the staff's insufficient rating.

MS. VIALL: Am I allowed to respond to that? And if my colleague, Irene Hall, is on the phone, I might also defer to her. But I think I will be honest that algorithms that we had to sort of walk through to submit this measure, they're very well suited for clinical quality measures. Whether they work for a measure that's based on public health surveillance systems or whether the application process is optimized for these kinds of measures, I think CDC would say it may not be. And so the algorithm we don't feel fully reflects the reliability and validity of our surveillance systems.

Anecdotally, I can say that we have an affinity group that we're working on with CMS and HRSA where state Medicaid programs are actually

clamoring to use the surveillance system viral load data to evaluate the quality of their Medicaid programs because they don't have access to test results, and we do.

So within states, there is a general feeling that the systems are incredibly reliable and valid. And where they tend to err, it's often a bias in the downward direction. So we're certainly not inflating anybody's performance with this. It's a conservative bias across the surveillance system, and improving that is something that states are doing iteratively through data to care sorts of activities, where they're actually using their surveillance systems to reach out to people to verify they're in care and, if not, they're looking at how do we improve the data coming into our systems, how do we improve the strength and validity of our systems.

So I would say that these are continuously-evolving systems, and we continue to put a heavy investment into improving their reliability and validity.

And, Irene, are you on the phone?

DR. HALL: Yes, I'm here.

MS. VIALL: Do you want to address sort of the fit of the reliability and validity algorithm to our systems?

DR. HALL: Yes. Basically, it's less that it's a faith in the process where we think that because the data are electronically reported and then matched and uploaded, they should reflect the actual values. And I agree with the fact, the assessment that it would be a downward bias because our biggest concern usually is that we get all data, all tests completely reported. But we don't have any evidence of any particular bias in terms of it being, the completeness being a bias. It's just a matter of states implementing complete laboratory reporting.

DR. NISHIMI: I just want to make one comment before we go to Amir. I mean, not Amir, Arjun and then Steve. We have had other measures come through NQF based on CDC surveillance. You looked at one yesterday coming through the NHSN

system on the healthcare vaccination for personnel. So I just wanted to push back a little bit on the notion that the NQF algorithm and criteria don't fit surveillance systems because we have that measure and other measures.

During the TA, we recommended them looking at perhaps state audit data, if they were able to get that, so if some states can show that their inputs, you know, are reliable and valid, then we don't need to, you know, necessarily show the whole system. Obviously, that's preferred, but that would have been quantitative evidence for the committee. But based on our algorithm, merely citing state law or indicating quality control and quality monitoring isn't sufficient.

So I'll go to Arjun and then Steve, and then I thought I saw another hand.

MEMBER VENKATESH: I think one of your comments sort of gets at what I was thinking, which is, this is challenging, right, because the data source that they essentially had is about as close to the gold standard that we would use for

some other measure. And so if somebody wanted to make an electronic clinical quality measure of viral load suppression to be used at a facility, we'd say, okay, how does this compare to some gold standard. That gold standard would be something that is collected and structured in a way, and often that would be this.

So the only thing I can think of, if we try to put this in a format of an NQF measure for something like this, is to say, okay, the underlying gold standard for this is probably some sort of audit that I'm sure has been done where you could simply just say, yes, at these many sites, we looked at 10 charts, 20 charts, 30 charts, and I'm sure 95-100 percent of the time the transcription of the number, which was that viral load locally into the system, was correct.

And my guess is that this has been, this has been up and running for a while. There's a large audit process behind this. I think I can live with that and say that I probably trust that that's right and that these

are valid. But I guess in an ideal world, if we wanted to make it fit the form, that would be the data that would say, hey, it fits the form.

MS. VIALL: And I would say we understand that. And, again, it's also the balance, we are constantly struggling with the states will routinely come back and say you're asking for too much. So there is, we have to balance that fine line of what we look at and what we take their self-certification on.

Irene, I think you can speak to this, but I do think we kind of say you are going to do these quality improvement activities and certify that you've met these. And from a federal level, since we're basically doing quality assurance several strata down, we have to accept that that's good enough.

Irene, do you want to talk about that a little bit more?

DR. HALL: Yes, I'd like to say two things. One is that lab reporting is monitored, and we provide states with quality measures on

how to monitor their lab reporting in terms of volume.

The other piece about audits, generally we ask states to do re-abstracting studies, but, in this case, a person could go to multiple facilities and it's very hard to find all the multiple facilities, while, when you have complete laboratory reporting, you will get those data from wherever the patient goes.

We do have another system. It's called the Medical Monitoring Project, which is funded in select jurisdictions. But since they are now also sampling from case surveillance, we will get some information at least for those select jurisdictions to see whether we can make any assurances about completeness.

MEMBER TEUTSCH: So, again, full disclosure, I'm retired from CDC.

MS. VIALL: I know. I recognized your name. I think we have an award named after you.

MEMBER TEUTSCH: And a surveillance book. So I think what I wanted to say here was

that CDC actually has a formal process of evaluating surveillance systems, and they include many of the same kind of criteria that we talk about here, but they're oriented in a different way. And the goal of those evaluations is do these surveillance systems provide information that's useful and can drive action? And I would suggest to NQF that, as we look at some of these measures that come out of those, that we begin to look at the guidance that comes out of those evaluation standards which get at many of the issues I think that we actually care about because the point is that they should be useful at some level of action, in this case the state, right? And we know that, in many cases, surveillance systems have a significant amount of error. So I told you a Salmonella surveillance probably gets two percent of the Salmonella cases. That doesn't mean it's not usable, but there's a certain constancy in it, and you can use it. Obviously, HIV is much better than that.

But I would suggest that we have some

criteria, CDC uses them, and that we look at them. And I don't know specifically whether HIV has done them, and they probably have. There's been so much attention, you probably meet all the criteria anyway. But there are some things that we could look at that I think would help us and get us out of the clinical care conundrum of holding these metrics, to that kind of standard, public health standards.

MS. CRAWFORD: Voting is open for Measure 3086 on reliability. One is high; two, moderate; three, low; four, insufficient. And one more vote. 0 high, 7 moderate, 5 low, 3 insufficient. We're at 47 percent moderate, 33 percent low, 20 percent insufficient.

MS. MUNTHALI: So consensus was not reached, so we'll continue and hopefully resolve this at the post-comment call. So validity, I think we're ready.

MS. CRAWFORD: Voting is open for Measure Number 3086, validity. One, high; two, moderate; three, low; four, insufficient. We

have 0 high, 9 moderate, 3 low, 3 insufficient. 60 percent moderate, 20 percent low, 20 percent insufficient.

MS. MUNTHALI: So we're right at the margin. If we just had one percentage point more we would have been able to pass it, but consensus is not reached on this criterion, as well. We will continue to feasibility.

CO-CHAIR QASEEM: Any questions or discussion on feasibility? It's also rated insufficient by staff, so do you want to comment on it, or anyone?

DR. NISHIMI: The insufficient rating really was derived from the fact that only 27 states or now 33 states and the District of Columbia. So as a state-level measure, we didn't have all states, so feasibility was, at some level, insufficient, and trying to endorse a measure for accountability. But, obviously, the committee can feel that it's still feasible and usable.

MEMBER VENKATESH: I would think this

is highly feasible. There's plenty of measures we endorse that cannot be used or implemented by everybody who meets the level of measurement. There's a lot of hospital measures where many hospitals it's infeasible, where some it is. And so, to me, it's been up, it's running, it's working. I think it's feasible.

MS. VIALL: Can I respond to that? So just a quick note. We rapidly expect this to be in the 40s within a year or two.

The other thing is that CDC actually sees a state that cannot report on this. That is, in and of itself, a commentary on their sort of, their performance. States that have no results are not following what CDC recommends, which is that this is sort of the standard for HIV surveillance at this point.

So when we do our state progress reports, not having this value is actually something that has incentivized a number of states to really look at their laws and try to stand up these systems. So a lot of states

actually see the gap in performance that not being able to report on this measure represents as itself very informative.

CO-CHAIR QASEEM: Any further discussion? Let's vote, please.

MS. CRAWFORD: Vote on Measure 3086, feasibility. One, high; two, moderate; three, low; four, insufficient. One more vote. We have 5 high, 8 moderate, 2 low, 0 insufficient. 33 percent high, 53 percent moderate, 13 percent low, 0 percent insufficient.

CO-CHAIR QASEEM: So now usability and use, the final one. Any comments or discussion? Any comments from staff? Let's vote.

MS. CRAWFORD: Voting on Measure 3086, usability and use. One, high; two, moderate; three, low; four, insufficient information. Okay. Our results: 4 high, 10 moderate, 1 low, zero insufficient. 27 percent high, 67 percent moderate, 7 percent low.

DR. NISHIMI: Okay. We won't vote on overall suitability for endorsement because you

had -- two of the must-pass criteria were consensus not reached, so this will come back after the comment period to the Committee, and then you'll vote on overall suitability for endorsement once those comments have been received.

MS. VIALL: Do I get to exit this seat now?

DR. NISHIMI: Yes, you do.

MS. VIALL: It's a little bit daunting.

CO-CHAIR QASEEM: Thanks so much, Abigail, for coming. How about we take a quick break, folks? How about we -- is ten minutes enough? I think ten minutes is too long, right? So 11:15 let's just get back. Thanks.

(Whereupon, the above-entitled matter went off the record at 11:04 a.m. and resumed at 11:16 a.m.)

CO-CHAIR QASEEM: All right. So the next ones that we have on the agenda are the malnutrition measures. And I was just actually

talking to Arjun, and he has a valid point. We just need to switch the sequence around. I think we need to probably discuss screening, following by diagnosis, and then it is the same measure, or is it going to make a difference?

DR. NISHIMI: We have to first do this one.

CO-CHAIR QASEEM: We have to do that one, first?

DR. NISHIMI: Oh, wait.

CO-CHAIR QASEEM: It is the same people or no?

DR. NISHIMI: No.

CO-CHAIR QASEEM: It is the same people. It is the Academy of Nutrition & Dietetics.

DR. NISHIMI: Yes, that's why we did it. We put this --

CO-CHAIR QASEEM: Okay, so that is the issue. All right. Well, will you be okay with that? It says there is an issue with simplification -- simple measures involving

discussion.

DR. NISHIMI: Yes, that is why -- that is one of the reasons we did it that way.

CO-CHAIR QASEEM: So, let's stick with the sequence, and let me welcome our two guests. Do you mind introducing yourself, please?

MS. MCCAULEY: Good morning, and thank you for having us. My name is Sharon McCauley, and I am Senior Director of Quality Management at the Academy of Nutrition & Dietetics. And the Academy, just to give you a quick overview, is our world's largest organization of food and nutrition professionals, and we represent over 100,000 credentialed nutrition and dietetics practitioners. And we do strive to improve our nation's health through advancing the profession of dietetics, through research, education, and advocacy.

Our organization is part of a multi-stakeholder initiative that is focused on addressing malnutrition, which is a leading cause of morbidity and mortality among older adults.

As many as 20 percent to 50 percent of our patients are at risk for malnutrition or are malnourished at any time in a hospital admission. Patients malnourished during their hospital stay have a greater risk of complications, readmissions, length of stay, all outcomes associated with increased healthcare costs.

Today, I am representing the measure steward of these four measures, and they are very focused on, of course, malnutrition and I am joined with Dana Buelsing, our Manager of Quality Standards and our Chief Science Officer, Dr. Alison Steiber, is on the telephone line.

Additionally, our measure developer, our partner, Avalere Health is with us. And today we have Joe Lynch, Director, and Angel Valladares, who is a manager, and they are on the Avalere's Evidence, Translation, and Implementation Practice Team.

The Academy worked in partnership with Avalere Health to develop this set of four electronic clinical quality measures or

eMeasures. We also have hybrid eMeasures, and we addressed the recommended care process for malnutrition. And we appreciate the opportunity to present for your consideration the four quality measures aligned with the evidence-based nutrition care process.

These measures were developed through multi-stakeholder consensus fostered from two national dialogues. We had a dialogue session back in November of 2013, as well as September of 2014, and they included representative from CMS, ONC, health plans, health systems, as well as providers and patients. And what we did was we prioritized addressing malnutrition to reduce risk of adverse outcomes.

The emphasis is particularly on the hospitalized elderly patients, and they are, again, showing evidence demonstrating a rate of malnutrition as high 38.7 percent. The testing and development of these measures represents the first step in the National Quality Improvement Initiative for Malnutrition, focusing on

forthcoming broad dissemination from measure adoption and implementation.

These four measures address key components of the recommended malnutrition clinical workflow. And this is how we conduct it in the hospitals. It focuses on the first four of a six-step process, beginning with screening of patients upon admission, then completing a nutrition assessment for those who were found to be at-risk, and finally, the development and implementation of a nutrition care plan for patients properly diagnosed with malnutrition.

To reach such a goal, our organizations have developed these individual hospital-level measures focused on malnutrition, and, again, we align them with a multi-step process to address malnutrition. These measures address a need to encourage proper management of the elderly patient population in the hospital. Eventually, evidence from implementation of the suite of performance measures can inform such a global malnutrition score.

Due to the nature of these measures and their alignment with the nutrition care process and our clinical workflow in the hospital, I just wanted to make sure that we inform the committee, and I know that it has already just been mentioned that our intention for the four measures follows the nutrition care process. So, therefore, the order of the measures would be a screening assessment, and then the intervention plan of care, and then the malnutrition diagnosis.

So, with that, I would just like to thank you again for the opportunity to introduce these measures to the standing committee, and we look forward to your questions that you have on these measures.

CO-CHAIR QASEEM: Thank you so much. So, let's start off with our first measure. It is Measure 3090: Appropriate Documentation of Malnutrition Diagnosis. And let's go with our leads, Amy and Jacki. Go ahead.

MEMBER SPANGLER: Can I ask a question

real quick to the developers?

CO-CHAIR QASEEM: Sure.

MEMBER SPANGLER: Did you guys consider putting all four of these into a composite measure? And if not, why not?

MR. VALLADARES: Great. Hi, everyone, my name is Angel Valladares. Yes, that is something that we actually have been discussing in our engagement with the multi-stakeholder group that Sharon mentioned and also with CMS. So, it was actually something that CMS brought to our attention as something that they would want to sort of pursue in the future. However, of course, in our research and in the development of these measures, we needed to first implement and develop individually performing measures that we can ensure that all individual measures have a performance score. And then once they are implemented over time, and we can sort of do more -- generate more evidence around those four measures and how they relate to each other, the goal is to develop that global malnutrition

score, where we could then have the users of the measures monitor on sort of the performance of the entire nutrition care process.

So, I think that is the eventual goal.

MS. MCCAULEY: And our concentration is that hospital inpatient stay for that elderly population.

CO-CHAIR QASEEM: So, Matt, on this, I would rather not get into the discussion of composite measure. Let's just go over it measure by measure because that is not what we have in front of us today.

So, Measure 3090, Amy or Jacki, you guys would like to open?

MEMBER AUERBACH: I should note: This is John. Again, this is not about that but just a clarification.

CO-CHAIR QASEEM: John, I would really like to interrupt -- I am so sorry -- because we are running behind on the agenda. So, we would like to move on with the individual measure review.

Amy, Jacki?

MEMBER MINNICH: Sure, and Jacki, feel free to entertain any additional conversation.

So, first of all, I want to make a clear distinction because a lot of the detail that we are going to be talking about in this measure there is a distinction between the documentation of a diagnosis of malnutrition versus nutritional screening assessment intervention. And so that applied to a couple of the different pieces of this measure in specific.

So relative to the evidence, that really came through. There was limited detail relative to that documentation process. We certainly recognize it is an important measure but, based on the evidence that we have presented, there was a concern that there was insufficient information to make that determination.

Jacki.

MEMBER MOLINE: I second that completely. I was waiting to see -- you

mentioned that there was a malnutrition score that you are hoping to develop but there wasn't -- it seemed like the measure was begging to have that as what was going to be documented. So, it was unclear to me how people could be judged on a nebulous measure. And to me, it was nebulous in that it wasn't clearly specified what you meant by malnutrition in terms of how it would be captured. And that led to some questions that we had in the review.

MEMBER MINNICH: I think the other point is around disparities. There was really no reference to any type of disparities, other than what you have mentioned with the geriatric population.

CO-CHAIR QASEEM: Any other general comments, before we start going from -- Cathy?

MEMBER HILL: Yes, just as a general comment, as someone who is board certified in geriatrics, a nurse practitioner who works with these patients every day in the hospital and sits on readmission committees that look at what

brings our elders back into the hospital, this is a really important topic for us in Texas and it was when I was in Florida. So, I would encourage you to give real specific feedback to the measures developer because I can tell you that every day I put on diagnoses of, you know, underweight and especially in our stroke patients and our community-acquired pneumonia patients, which we have already talked about. These people are coming back because they are malnourished and our treatments, our medical treatments aren't working.

So, I really appreciate the fact that you have brought this to this group because it is an important wellness and health factor, nutrition.

MS. MCCAULEY: No, we have --

MEMBER MOLINE: I don't think that there is any doubt about the importance of this. I think as we were trying to go through and look at the measure, we were looking at what was presented to us. And what we were looking for

was some specificity; especially this one was looking as an eMeasure and was trying to figuring out what they were measuring.

So, that was really the issue, not the importance and its effect on morbidity and mortality.

MEMBER HILL: Yes, I appreciate that. And I was just trying to bring some personal experience to it so that you could hear that it is out there.

MEMBER MOLINE: It is clearly an important issue but the question in front of us is really more -- boy, I sound like I'm on that side of the table -- but the question in front of us is --

CO-CHAIR QASEEM: Are you an attorney?

MEMBER MOLINE: Yes, I spend too much time with them. The question really is what are we looking for to use as a measure.

CO-CHAIR QASEEM: Do you have a quick response?

MR. LYNCH: Sure. My name is Joe

Lynch. I am a Director with Avalere Health.

The primary thing that we are looking to measure is the actual medical diagnosis of malnutrition and the rate at which it is being done. If we look purely at how the medical diagnosis is documented, we have looked at it and saw the rate are down in the three to four percent of the actual ICD-9, ICD-10 documentation of diagnosis. But if you look at the literature and the studies around the rates of malnutrition, the estimation in the same time period is anywhere from 33 to 54 percent. So, there is a humongous gap in how effectively diagnosis is being documented in the patient record versus what is estimated to be the prevalence of the condition in the population at large, especially the 65 and older group.

So, the goal to this particular measure is to help to inform this concept of a global malnutrition score based on how effectively the clinical documentation of this diagnosis of malnutrition is taking place.

MS. MCCAULEY: And just to move backwards to explain a little bit more about the diagnosis and I understand -- thank you for your comments -- it is the end result of going through that process and that clinical workflow.

So, I understand what you are saying. You want that stand-alone measure to have I guess that cause and effect. But what we do is making sure that the screening, the assessment, and we work with a multi-disciplinary team to have that happen through that nutrition care plan to make sure that we work with a physician to get to that diagnosis. And that is how we approach this measure suite.

CO-CHAIR QASEEM: So, Marcel and then Tom.

MEMBER SALIVE: Though, I didn't hear us discuss this for pneumonia and I think -- I appreciate that malnutrition is underdiagnosed in the hospital and that that is the focus of this measure, I do think it is an important issue and it is worthy of our consideration. And so I am

not clear on why that is the focus on the evidence section here. You know we, generally, I think accept diagnosis codes that are in the records and that is what you are saying here.

So, you know maybe it is way vastly under-reported now but I do think that, as we discuss these other measures, we will get to that. But if you don't do screening and if people aren't looked at in the hospital, it won't be picked up. And I think it is worthy of doing and it is a point of starting the process, when it could happen.

People are, essentially in the hospital. They have to eat there. I mean unless they are unable to eat, their nutritional needs have to be met. So, it seems reasonable to me as a focus for a measure. And I think there was some evidence that was presented that is sufficient for this to be a measure. Although, personally, I think, some of the other measures might be better. But I don't understand -- maybe the reviewers can explain this better but I don't

understand why that is your issue. We didn't -- I guess we discussed with pneumonia that some are viral, some are bacterial, but we didn't say people can't diagnose it.

CO-CHAIR QASEEM: Amy, Jacki, you want to comment? Since we are getting into the evidence piece anyways, we can talk about evidence even from the staff aspect of it.

DR. NISHIMI: Yes, I mean there is evidence for each measure is what I heard the two reviewers say and that the evidence that was cited is more relevant to the other measures, not to this measure. I don't want to put words in your mouth but I think that is what they are saying.

So, we'll get to those but the measure right now is just the documentation.

CO-CHAIR QASEEM: Matt?

MEMBER STIEFEL: I apologize. I can't not talk about the combination because we are going to fall into the same trap as we did before, that the evidence is insufficient for

each of the independent measures but the evidence may more appropriately apply to the bundle.

So, I have a feeling we are going to go through this process of insufficient evidence in each of the components.

CO-CHAIR QASEEM: John, why don't you chime in? This will be a good opportunity since Matt brought it up.

MS. MCCAULEY: I don't think it will--

CO-CHAIR QASEEM: Matt brought it up again. John, do you want to say a few words about the whole composite measure issue? You were planning to chime in earlier.

MEMBER AUERBACH: Thank you very much. The question I had, which I think applies to all but, certainly, we can limit it to this measure for now, is whether or not you are making a recommendation that the screening occur for elderly patients or for all patients above 18 years and older.

Partly, this is a question related to the evidence, since part of what the evidence --

part of the submission of the evidence focuses on the elderly and your comments have focused on the elderly.

So, could you clarify whether this is a more of -- you know what the focus of the population and perhaps whether or not that has an impact on the availability of evidence?

MS. MCCAULEY: We decided that the screening tool that we were using will be for all patients, 18 and above. So, that is our first measure that we had worked on.

And moving forward, we isolated it to the inpatient elderly hospitalized.

DR. NISHIMI: Just to address Matt's concern about insufficient, I don't think you will reach that conclusion, just to preview the other measures and the reviews that folks have laid out. I don't think the committee will reach an insufficient conclusion. This was the measure about which I think people are likely to --

MEMBER STIEFEL: But then the problem is that the other measures, without this one, are

insufficient, to achieve the outcome goal.

MEMBER MOLINE: No, one is a screening for it and then the other is to put in the record as a diagnosis.

So, I think they are actually, one is looking did you do the screening and I think that is what we will be hearing about later.

The other is you did the screening. Did someone pay attention to the screening and actually add that as one of the diagnoses that would then be acted on to improve the quality?

So, they are actually -- the screening needs to come first. Is it happening? Because you need the screening in order to get the diagnosis but I think this measure is saying once you get the diagnosis, is it actually making it into -- is it actually getting truly documented and is there evidence that there is ease of doing that and how it is being done?

DR. STEIBER: Hi, this is Alison Steiber. I am the Chief Science Officer for the Academy and I would love to just briefly address

that point.

So, I apologize if I am a little unclear of the process here.

So, we have pretty good evidence from a long history of JCAHO requiring screening in the hospital within the first week of admission --

DR. NISHIMI: If you could hold your -- I'm sorry to interrupt you but the committee is discussing this measure now. If you could hold you comment on screening, when we get to that screening measure.

DR. STEIBER: Okay.

DR. NISHIMI: We are just talking about the document -- evidence around documentation per se right now.

DR. STEIBER: Right. So, I guess I was just going to comment on the disconnect between the screening happening and the diagnosis occurring. Is that appropriate to speak on at this moment?

DR. STEIBER: It is appropriate to

comment on the documentation of the diagnosis because that is what this is; not the diagnosis.

DR. STEIBER: Right. Okay, so my point was just simply that we have evidence through a number of surveys that were conducted that there is a disconnect between people who are being screened at risk for malnutrition and the documentation of the diagnosis of malnutrition and yet while we have that significant disconnect in that documentation of diagnosis of malnutrition, there is evidence to indicate that we can successfully diagnose malnutrition and when that is done, we have a difference in survival rates and in costs in the hospitalized patient population.

CO-CHAIR QASEEM: Cathy.

MEMBER HILL: Yes, and I agree that what I -- just from a practical standpoint, what we see happening is that our -- I am most involved with elders because that is 50 percent of my inpatient population in rural Texas. We do see them getting haphazardly screened and our

estimates in our readmission efforts have been anywhere from 13 to 78 percent of our population have needs that are not being met.

The interventions are been getting done separate and apart from the diagnosis, which has been problematic in terms of driving a consistent process that improves the health of our patients and reduces our readmission rate. So, the diagnosis is a part of, I think as Matt had suggested, is a part of how you get it done is establishing that label.

CO-CHAIR QASEEM: So, I think we can vote on the evidence piece of this, at this point.

MS. CRAWFORD: Voting on Measure 3090, evidence. One, high; two, moderate; three, low; four, insufficient.

MEMBER BAER: Hi, this is Mike Baer. I just wanted to let you know that I did join. I was unable to join earlier but I am here now and I did vote via the chat.

DR. NISHIMI: Thanks, Mike.

MS. CRAWFORD: Zero high; five moderate; four low; seven insufficient. 31 percent moderate, 25 percent low, 44 percent insufficient.

DR. NISHIMI: So, the measure does not pass the must-pass criterion of evidence. I am going to move on to the next measure.

CO-CHAIR QASEEM: Okay, so the next measure is Measure 308 -- actually, can I just change the sequence around this time around? Because I think it would make more sense if we discuss the screening first --

DR. NISHIMI: Sure.

CO-CHAIR QASEEM: -- which is measure 3087. And I think it's the same reviewers, correct? I think it is Jacki and Amy. Am I right or am I wrong? Hold on, I have a cheat sheet I have to look at.

DR. NISHIMI: 3087.

CO-CHAIR QASEEM: Sorry to move things around. Sorry, guys but I think it will make more logical sequence.

Okay, so it is Cathy and Barry is not there and Ron. So, Cathy and Ron. Who would like to --

MEMBER BIALEK: I'll chime in initially, okay?

So, this measure is to have screening done within 24 hours of an individual being admitted to the hospital. And the progression is screening and then it is screening would identify high risk. High risk then would result in a more complete assessment, which then could go into diagnosis, treatment, et cetera.

And as far as the evidence is concerned, the evidence presented demonstrates that there are adverse health outcomes associated with malnutrition and that the intent of the measure, again, is to start the process of screening to assure that, ultimately, it can lead to the diagnosis and treatment.

So, that is the evidence piece.

MEMBER HILL: I agree. I don't have anything to add. I think the evidence is there

and my experience over the last 30 years would support that.

DR. NISHIMI: You can see from the staff PA -- can you call that up -- that there was a systematic review of the evidence, the quality, quantity and consistency. It has provided the evidence was graded. Depending on the particular sub-recommendation, if you will the grades and level of evidence were different but there was grading provided. So, the committee can vote. It is eligible for high and the committee can vote on evidence.

CO-CHAIR QASEEM: So, before we vote, Ron, Cathy, a question for you. And please, I reviewed this measure a while ago and I have some notes over here and I was looking at it.

My question, and I am not disagreeing with you, Cathy, what you are saying is that this is talking about screening all patients over 18 years, rather than just ICU patients, older, elderly, right, the population that you were talking about?

And one of the notes I have from the measure reviewers, if it is in the measure somewhere, I have it in quotes, which says in the additional comments included below, the measure developer would like to acknowledge that it is difficult to measure patient outcomes in the nutrition space, particularly to associate outcomes with only one of the steps of nutrition care in the overall nutrition care process, et. al.

So, my concern is that maybe I am forgetting it but, Ron, it is all patients. Did they present evidence for all patients in there or did I miss it? I understand the ICU patients. I understand the elderly and all. You are talking about everyone over the age of 18. Are you looking at almost, quite a big chunk of U.S. population who is going to start getting this screening done?

MEMBER HILL: Well, my additional focus is in rural health, where socioeconomic issues abound and I frequently have patients who

say they don't have access to good water --

CO-CHAIR QASEEM: Absolutely.

MEMBER HILL: -- and they don't have access to food.

CO-CHAIR QASEEM: Yes. So, I mean it is a specific population. I'm not disagreeing on specific population at all. I am talking about they are talking about screening everyone.

Is there evidence for screening everyone, every adult person in the U.S.?

DR. STEIBER: This is Alison Steiber again. Just to indicate that the data that is out there on screening and prevalence of risk for malnutrition is not subjected to just 65 and over. Most of the data that we have globally includes adults 18 and over and so that 33 to 54 percent that Joe quoted before is from a population that is general adult, not just for elderly.

MR. VALLADARES: And this is Angel. I just wanted to add that the guideline that was cited for support for this measure does recommend

screening for nutrition risk for all hospitalized patients. So, that is something we wanted to add. It doesn't distinguish between age ranges.

CO-CHAIR QASEEM: And what was the rating for the evidence? That is exactly the concern that I am raising.

MR. VALLADARES: Right. Great. And so I think that goes back to the comment that we had made, that you brought up in the additional comments for the measure submission. So, one of the challenges that we have in the nutrition space with research is sort of the inherent way that a lot of the research is designed. There isn't a lot of randomized control trials, as you can imagine on malnutrition and screening. So, a lot of the evidence, unfortunately, stays in the traditional research space more on the level, I think it is Level 3, and 4, and 5 as we define it.

So, a lot of it is on observational studies, cohort studies, retrospective reviews, as opposed to pro-RCT type of research. And so I

think that is where the issue is. But in terms of the volume of research, it goes back decades, confirming the prevalence hasn't changed. In fact, it is getting worse as our population ages. So, it is a continued issue but I do understand it.

CO-CHAIR QASEEM: Absolutely. And I will come to Arjun and Cathy just to respond to that. And disclosure: I am a member of, being a member of the Grade Working Group. Just because you don't have evidence from a randomized controlled trial does not make it a low quality evidence. I think that is something that we need to keep it in mind.

If you are unable to derive practical recommendations based on randomized controlled trials, you can still upgrade observational studies as well. So, I am going to a little bit disagree with that.

I think the reason that was rated as Grade E is because it is Grade E for various reasons that are in that guideline. That is why

I was bringing up that issue.

Arjun and then Cathy.

MEMBER VENKATESH: So, I guess from an evidence perspective, I am hearing that this is probably again down that path of insufficient evidence and then you have to decide if you want to make an exception or not based on international or national consensus statement. I think when you make a recommendation for a quality measure evidence based on a specialty society consensus statement, we need to look through and make sure that that consensus statement grading was appropriate and things like that. And I haven't looked at that detailed a measure.

My issue with this, also sort of related to that, issue about the evidence is that the measure itself is set up so that -- sorry -- I'm trying to find it -- if we were to do this, it is not innocuous to do this on very low levels of evidence.

There is something probably around 25

or 30 million inpatient hospital discharges -- inpatient. This measure is not specific to inpatient observations. So, if you included both, people who stayed in the hospital 24 hours, you are looking at something on the order of 35 million, let's say, times a year, where you would have to screen people and this would include things such as somebody comes in for elective gallbladder surgery would now need to get a malnutrition screening. Every single hospitalization.

This is a very broad screening measure. And to me, I think there needs to be some level of evidence that suggests that our pretest probability for malnutrition, our suspicion of malnutrition, you know screening is smart. We wouldn't accept the same level of standard for screening. If somebody just came in with a colonoscopy measure and said we should do a colonoscopy on everybody above the age of 18, we would say no.

And so I think that -- I do think it

really matters, at least on the age or some other risk elements of a patient. I understand there are many people under 65 who are certainly at high risk and there are probably ways to define that risk cohort, based on observational research. You don't need randomized trials. But I think a broad measure of everybody above the age of 18, 30 or 40 million times a year in the U.S. just seems a little excessive.

MEMBER HILL: Well, tangentially, I think, related is the fact that in population health and in outpatient settings, we often feel like it is appropriate to talk to people about their nutrition. And we all embrace the role that nutrition has in staying healthy and well. Many patients have a disconnect, as do providers, that all of a sudden you come into an acute care setting. We don't really care to make that connection between what is going on with your hospital stay and your ability to get well successfully and your nutrition and how that plays a role. And we are at the beginning, I

think, of trying to do that, to harmonize with the message that we give to populations about lifestyle, and the 30 percent role it plays in how well you are, and connecting that in the acute care setting, so that that remains a relevant part of the role and gets reinforced by people who are respected and trusted, like physicians and nurses and dieticians and nutritionists.

DR. STEIBER: Sorry, this Alison Steiber one more time. I couldn't --

DR. NISHIMI: Alison, can you let the committee --

DR. STEIBER: Oh, sorry. Sorry.

MEMBER SALIVE: In reviewing the measure, what I didn't see is the extent of the initial screening. What I took away from the description is that it wasn't all that extensive in that the more extensive is the malnutrition assessment that will be prompted by this initial screening that didn't seem to be all that time consuming. But that wasn't explicitly said in

there.

Secondly, the other thing I would say is that the evidence presented does show that use of a validated screening tool does identify both within the elderly and the general population more high risk that leads to the full assessment, that leads to diagnosis of malnutrition. So, that was presented, I believe, in the evidence.

But the measure itself, screening, isn't specific to use of the validated screening tools. While that is recommended, that is not specific in here. So, it was tough to really gauge all of the evidential pieces in that using screening in general, it is not validated. I would say that there was not evidence for using just a general tool that is not validated but there was evidence for using the validated tool but the measure doesn't specify it has to be a validated tool.

MEMBER SALIVE: So, maybe this is out of order. I agree with Ron a little bit but I think this is -- the sure tools that I am

familiar with and that are cited in the guideline include BMI, which we will get to in one of the other measures, which is why I say it is out of order, and then it just has like four questions.

So, this is not like super complicated and it is basically have you been eating less in the last three months. I mean that is very simple. People can answer that. And have you lost weight in the last three months?

And then the other two are related to more like physical and mental illness. So, I think that is known on the hospital admission. This is not super complicated. I think that people do need to be fed. This is not burdensome.

So, yes, okay, if you don't say it is a validated instrument and maybe there is some wiggle room but that instrument is very simple. And we will get to the one on BMI, which I think has strong evidence and strong recommendations. So, that is a component of this screening and it is probably the most important component, I would

guess.

CO-CHAIR QASEEM: Okay, so Alison and then the team.

MR. LYNCH: This is Joe. Actually, first to address the burden here. Up until just the beginning of this year, the Joint Commission required that malnutrition screening occur within 24 hours of admission. The issue with that is there was no measure associated with collecting information about that screening. So, it was a check the box process with Joint Commission accreditation. So, this is not new. The idea of doing this isn't new. It is part of the process already.

The validated screening tools, though important, and you are absolutely correct, the evidence does support using the validated screening tool more specifically, the fundamental issue we ran into is the ability to actually capture that bit of information in the EHR. So, the ability for us to verify the existence of a validated screening tool and the measure was a

challenge, at least at this point. Measurement can change that behavior, ultimately, but as the data exists today, that is difficult to do.

And to answer your question about the complexity of the tool, it is usually about three questions. It is a very simple screening process to get it done.

But in terms of to address the overall issues, it is a very simple process to get done and is shown to really drive how effectively further evaluation of a person's care and the existence of malnutrition is a sensitive enough tool to identify those at risk for or usually at high risk for malnutrition. The middling ranges in the potential for risk will start to bubble up to the surface but that is not the intent of the tool. It is really to identify where things can be addressed, where the evidence really supports that intervening, at this point, does result in much improved outcomes.

Did you have anything else more you want to add?

CO-CHAIR QASEEM: Alison, do you want to add anything?

DR. STEIBER: You know I guess I was just going to point out that one of the challenges, even from the JCAHO time when they were mandating this is that every hospital had their own concept of what was needed for a screening tool. And I think the original intent was really to use a validated tool but the question of feasibility came into play whether that was feasible for one tool to meet the needs of every hospital, whether it is a community-based hospital or a tertiary medical center. And so I think that it was felt that it was more feasible to have a little bit more open ability for the hospitals to have tools that they felt would be effective for them to identify risk of malnutrition.

And clearly, as Joe nicely put it, the goal is to improve outcomes. And so we have to figure out how to support the facilities to do that in ways that are feasible for them.

CO-CHAIR QASEEM: So, I think we are probably ready to vote on the evidence. Marcel, do you have a comment? Oh, okay. Let's vote, folks.

MS. CRAWFORD: Voting is open on Measure 3087, evidence. One high; two moderate; three low; four insufficient.

One more. And we have eight moderate; two low; six insufficient. So, 50 percent moderate, 13 percent low, and 38 percent insufficient.

DR. NISHIMI: So, that is consensus not reached and we will keep discussing this.

CO-CHAIR QASEEM: Okay, so let's keep discussing it.

DR. NISHIMI: We will discuss the other criteria, I should say, not that we are going to continue discussing this.

CO-CHAIR QASEEM: You want to get to the gap part now, right?

DR. NISHIMI: Right.

CO-CHAIR QASEEM: Okay.

MEMBER BIALEK: So, a performance gap was demonstrated but what I wanted to point out is that the data come from only two hospitals and there was no information on generalizability.

DR. NISHIMI: So, that is something for the committee to consider in its deliberations.

CO-CHAIR QASEEM: Cathy, do you have any comments?

Any general comments no gap? Okay, so let's vote on gap.

MS. CRAWFORD: Voting on Measure 3087 on performance gap. One, high; two, moderate; three, low; four, insufficient.

We have four high, nine moderate, two low, one insufficient. So, 25 percent high, 56 percent moderate, 13 percent low, 6 percent insufficient.

DR. NISHIMI: So, the measure passes on gap and we will go to the scientific acceptability. That is the reliability of the specifications, the reliability testing and the

validity testing and assessments of threats to validity.

So, Ron and Catherine.

MEMBER BIALEK: In terms of reliability and validity and testing, the measure itself, again, was for screening. And the determination of if somebody has been screened really seems to be up to the provider, him or herself, the coder, him or herself. There really was no specificity on what is screening.

And the evidence, again, about using a valid tool was clear but you could code it as somebody being screened without using that tool or any standard process. And so I found some difficulty in the reliability of the specification of the data, as well as the comparability as a performance measure that screening is not the same as screening, depending upon the provider, the institution, et cetera.

CO-CHAIR QASEEM: Cathy, any --

MEMBER HILL: Well, there is something to be said for the fact that this has been

suggested for the inpatient setting. So, you are going to be screened by licensed, credentialed people and while that may not be, at this point, may not be standardized across the nation with the same four questions or three data points, you are getting a level of expertise there that is appropriate to the acute setting.

CO-CHAIR QASEEM: Any other comments? So, I have a combination of comments. This is reliability, validity and the EHR altogether.

So, there are no exclusions for this measure, my understanding is, and which surprised me a little bit because when you conducted the EMR feasibility yourself, you excluded patients. You excluded patients discharged before 24 hours, discharged before hospice, and then two or three other categories in there.

So, what confuses me is that you don't have an exclusion but when you tested the measure, you had exclusions yourself.

MR. VALLADARES: That is a great point. So, I think before we conducted the

actual validity and reliability testing, we did test for feasibility on those particular sub-populations. Once we did testing, and you will see that in the reliability and validity testing that we ran tests on both patient populations, excluding those patients and patient populations not excluding. So, we did run a pretty specific measure exclusions analysis on both of the testing sites and the cohorts. We did not find that the exclusion of -- first of all, the exclusion and the number of patients that were excluded were very small. It was a three to five percent, if I recall.

And then, in addition, once we looked at whether it would impact the data element results or the measure performance results, it had no significant impact at either level. So, I think that that is the reason we wanted to present the way we had the feasibility first tested but then we showed in the measure exclusions analysis that our initial indications for excluding those patients, which was supported

by, you know they were things that were brought to consideration to us by the technical expert panel. And then once we presented that information back to the technical expert panel and we had the reliability and the validity testing scores, they agreed that because it had no actual impact that we could assess through statistical analysis. we shouldn't include them until such time in the future where there are larger amounts of hospitals that might change the evidence.

MEMBER VENKATESH: So, I guess this is maybe more guidance on this issues, in some ways, which is that the vast majority of exclusions for most quality measures will not be captured in any meaningful or high rate. And for example here, you only looked at 200 records. I'm not surprised that these three things happened at very low, if did not happen at all.

I would say that my interpretation of this same stuff that you have presented here is that you had a technical expert value that said

the face validity of your measure requires that you exclude these three populations because it would be unreasonable to expect that they get screened within 24 hours. You evaluated whether or not those can be feasibility captured. I think they all can because they are actually part of other eMeasures that exist. Look at the ED throughput measures include all of these to some degree in electronic specifications. So, they can feasibly be captured.

Looking at scores with and without exclusion had a high degree of correlation. So, there is not something systematic about exclusions that throw off your scores but you can still keep them in as denominator exclusions because they improve the likely meaningfulness of the measure.

You could imagine that if you eventually roll this out on every hospitalization, thousands and thousands, you are certainly going to have all three of these things happening with some different degree between

different facilities. Hospitals have different rates of referral to hospice. They are going to have a different amount of short stays. They are going to have different amounts of patients that leave AMA.

And so for all the reasons that your experts said make them exclusions, I would make them exclusions. I guess I would come to a different conclusion with the exact same data findings you have here.

CO-CHAIR QASEEM: Okay, any other comments? And then we can vote.

MEMBER HILL: I would like to agree with that and I do know that when you are trying to -- this is a trial measure -- no?

DR. NISHIMI: This is not a trial use measure. This is a measure for endorsement.

MEMBER HILL: All right. Then, the adoption of measures is definitely influenced by the exclusions seeming reasonable to the level that you are going to implement it.

CO-CHAIR QASEEM: Okay, so let's vote,

please.

MS. CRAWFORD: Voting on Measure 3087, reliability. One, high; two, moderate; three, low; four, insufficient.

DR. NISHIMI: Re-click.

MS. CRAWFORD: One more. Zero high, 11 moderate, 2 low, 3 insufficient. So, 69 percent moderate, 13 percent low, 19 percent insufficient.

DR. NISHIMI: So, we can continue and discuss validity, which is where the exclusions discussion really belongs.

CO-CHAIR QASEEM: So, any further -- go ahead, Ron. Shall we vote?

DR. NISHIMI: Is everyone comfortable with no additional discussion? We can vote, otherwise.

CO-CHAIR QASEEM: I think we can vote. Let's vote because we have combined the discussion last time.

MS. CRAWFORD: Okay, voting is open for Measure 3087, validity. One, high; two,

moderate; three, low; four, insufficient.

One more vote.

We have one high, nine moderate, five low, one insufficient. So, 6 percent high, 56 percent moderate, 31 percent low, 6 percent insufficient.

DR. NISHIMI: That is 62 percent high or moderate. So, it passes on validity.

CO-CHAIR QASEEM: Okay, so feasibility, Cathy and Ron?

MEMBER BIALEK: Okay, feasibility. So, this can be publicly reported. So, did you just say feasibility or usability?

CO-CHAIR QASEEM: Feasibility.

MEMBER BIALEK: Feasibility. I'm sorry, feasibility. I actually thought there was insufficient information to determine this across hospitals just because there were two hospitals in the sample and, again, no information on generalizability. So, I really couldn't weigh in on that.

CO-CHAIR QASEEM: Go ahead.

MR. VALLADARES: So, in response to that comment, one thing we did want to clarify because I feel that it hasn't been included in the discussion on feasibility is we actually tested on three hospitals, feasibility and we also tested on three national EHR vendor platforms.

So, we worked with Epic Systems, Cerner, as well as Allscripts, who are three of the largest EHR vendors for hospitals and they all provided sufficiently, well we believe, at least, above average feasibility, saying that they were able to capture this data electronically in their systems and they felt that the coding and value sets that we included to categorize the components of this measure were absolutely capturable to their customers.

And just as sort of a quantitative assessment, if you aggregate the hospital platforms that those three vendors are responsible for represents a little over 30 percent of the U.S. hospital market who have 2014

certified EHR technology.

CO-CHAIR QASEEM: Go ahead.

MEMBER BAER: So, can this be generalized to all certified EHRs?

MR. VALLADARES: That's a great question. So, yes, these data elements represent value sets that follow the healthcare quality measure format and they are all approved in the VSAC so they are, again, up to the national standards. Our three EHR vendors who we work with confirmed that the measures -- or sorry -- that the data elements and the value sets included in those data elements represent a nationally standardized data that is already implemented in their platform, in their suite of platforms that they provide to their hospitals.

CO-CHAIR QASEEM: Katie.

MEMBER SELLERS: Yes, I just had a clarifying question for NQF. I believe the threshold requirement is to test it in two sites, right? So, they are well above that threshold.

DR. NISHIMI: Two systems, yes.

MEMBER SELLERS: Systems.

DR. NISHIMI: Two systems.

MEMBER SELLERS: Yes.

CO-CHAIR QASEEM: Except, just so to follow-up, Katie, on what you just said, to my knowledge and, again, I am not the expert in this field, you are talking about screening everyone over the age of 18 and most hospitals do not have enough dieticians to be able to consult among all patients who might be needing this. And I don't know if your hospitals, the three hospitals that you included. Where did they fall into?

Because the lack of dieticians -- you screen if you are going to be able to provide certain services and adequate treatment. If you are not going to have the follow-up in place, you do not screen. That is the basic rule for any screening recommendation. So, these three hospitals that you guys had, the dietician to patient ratio or if you can speak about --

DR. STEIBER: Is it appropriate for me to comment on that? This is Alison.

CO-CHAIR QASEEM: All right. Go ahead, Alison.

DR. STEIBER: Great, thank you. So just to clarify, in most hospitals, as soon as -- it is not the registered dietician/nutritionist who does the screening. Screening usually is done by either dietary technician registered or often by the nursing staff. And so it is not until the assessment step occurs that the dietician would step in.

So, I certainly agree with your point on staffing issues but the screening step really typically is not done by your registered dietician. So, the staffing typically has not been an issue, at least when JCAHO required it.

CO-CHAIR QASEEM: No, absolutely. I agree with you that the assessment is going to be done by nursing. What if you need to get dieticians for consulting for any of these? It is not going to be?

DR. STEIBER: Yes, -- no. So, the assessment, actually I would argue, is

appropriately done by the dietician. It is the screening step that I believe is typically not done by the dietician. So, screening, then assessment, and when they are at risk for malnutrition, then they hand it over to a dietician and that is when that step occurs.

And I do believe that we do have sufficient registered dietician staffing in most medical centers to handle the patients that are screened as at-risk. However, certainly this measure may shape that staffing ratio even further.

CO-CHAIR QASEEM: And I think we can -- if there are no other comments, I have a really dumb question for you all. You are more the experts than I am. How do you enter this in an EHR in Epic, the screening, this information? I am trying to figure out how will I enter it and how will it get extracted.

MEMBER HILL: I can tell you how it is done in my environment. And that is, it is a checkbox on the admission assessment that

triggers an automatic referral to the nutrition dietary department for them to look at that. And then in our environment, the dietician will screen the chart and date it to see if it is truly worth the follow-up.

CO-CHAIR QASEEM: Because it was a feasibility issue, others who might have experience with this, how do you guys enter this in your Cerner or Epic? Because in the EHR as you write -- so, how are you going to enter the information and then be extracted for someone to be able to act on it?

MEMBER HILL: Well, I had experience in Epic.

CO-CHAIR QASEEM: Oh, just Epic in your hospital?

MEMBER HILL: I can't -- well, I have experience in Allscripts, Epic, and MEDITECH. And in all three of those --

CO-CHAIR QASEEM: It is the same?

MEMBER HILL: Yes.

MEMBER VENKATESH: It is the same.

CO-CHAIR QASEEM: All right, Arjun same? All right. Sounds good.

Anything else?

MS. MCCAULEY: And just to follow-up, after that registered dietician then gets that information, there is further triage. So, it may go to very specialized dieticians versus generalists, versus technicians, as Alison mentioned. We refer back to nutrition assistance. So, we make sure that we really get to those patients that have been assessed and then we answer that consult.

CO-CHAIR QASEEM: Can we vote on feasibility?

MS. CRAWFORD: Voting on Measure 3087, feasibility. One, high; two, moderate; three, low; four, insufficient.

And we have two high, twelve moderate, two low, zero insufficient. So, 13 percent high, 75 percent moderate, 13 percent low.

DR. NISHIMI: Usability and use.

MEMBER BIALEK: Back to the issue of

getting the screening, not being specified for the use of the validated tool, which is, again, where all of the evidence seems to suggest is the appropriate approach, I would say in terms of accountability and performance improvement, this really could not be used that way because of the variability of the practice of screening in this instance.

CO-CHAIR QASEEM: Any other comments, Cathy?

MEMBER HILL: I would like to suggest that we consider this -- can we consider this measure as with one exception on the evidence? Because I think that the evidence will improve with --

DR. NISHIMI: We already passed it on evidence. So, you wouldn't go back.

MEMBER HILL: Okay.

DR. NISHIMI: You would have to decide -- you know basically, the final -- on the final vote will need to make their own decision about is it suitable for endorsement or not.

CO-CHAIR QASEEM: Okay, so seeing no other comments, how about we vote on usability and use, please?

Sure, go ahead.

MR. VALLADARES: So, one thing we did want to highlight and, of course, it is a little different under the purview, possibly, but the particular measure set that you are looking at today is actually part of a National Measurement Quality Improvement Initiative that is being led by multiple organizations, including the Academy of Nutrition and Dietetics. Right now, the actual strategy and plan for this initiative is to expand the use of the measures and implement a standardized malnutrition toolkit that actually recommends validated screening tools. Right now, there are about a subset of six to eight hospitals that are beginning to pilot not only the measures but also the toolkit. And the goal is to expand it further. So, that is just something to keep in mind, in terms of usability that there is an actual, since the implementation

when we first tested the measures, sort of the ball is moving forward with usability of this measure with the hospitals. I just wanted to share that.

DR. NISHIMI: Ron?

MEMBER BIALEK: I absolutely agree with what you are saying. However, again, the measure is not to use the validated tool. So, you know the suggestion is that usability is being demonstrated through these other means with using the validated tool but that is not what this measure is for.

The measure is for screening not for validated screening.

MS. MCCAULEY: And just to narrow that, we do have several facilities across the country that do use validated nutrition screening tool in their whole health systems. We determine, and that is why we have several studies that justify that use. I think when you have the complexity and the difference in the QE rate and the different conditions in the

hospital, we allow nursing and the dieticians with the doctors and the pharmacists in speech/language in that multidisciplinary team that they are working with to determine what those questions are going to be on that admission item and what those checked boxes are.

Many times, because of the specialization, there is core business at that hospital, the dieticians may add a few more, the nurses may want to have a few more questions. Hence, that is why then they may take some of the first two or three questions from the validated score and tool but they will add a couple more questions, just to make sure that they are capturing what they need for those patients who could be at risk in those certain conditions.

DR. NISHIMI: Emilio.

MEMBER CARRILLO: Yes, just kind of a point of information in the question. CMS is now spending a lot of time putting together a screening for food and security as part of the Accountable Health Communities Program. Is this

in sync with the work that you are doing or is there any connection?

MR. VALLADARES: So, and I will let Sharon actually follow-up because I am not 100 percent sure but I know that the work that we are doing with malnutrition in general in a hospital is very focused on the inpatient setting with CMS, at this point. But the intention in the next phase of the project is actually to expand outside of the hospital care setting, since we do know that elderly malnutrition is a community-level issue. But this is sort of the start of the rollout across the care setting continuum.

The other thing I did want to add, if I may, if it is possible -- I'm not sure.

So, the other thing, this is just back in -- and I want to put this into context with the Joint Commission because I think this is really important. So, the actual reasoning the Joint Commission removed the standard of nutrition screening within 24 hours from its clinical standards is because they felt that it

was such a standardized process across all of their accredited hospitals across the nation that they felt that they could focus on something else. However, the reason why we are bringing this measure as a quality measure and focused on accountability is the fact that, one, they removed the standards. So, of course, now the hospitals don't need to focus on it. So, we are afraid -- so, there is always that understanding.

And then in the evidence attachment that we presented, we showcase that study after study proved that the actual risk, rate of malnutrition, and then the eventual assessment and findings of malnutrition don't balance out with what is in terms of surveillance when you look at ICD codes only. So, only claims assessments of population surveillance, that number is very, very highly underreported. And so there are millions of cases of malnutrition that are being left undiagnosed because there is no systematic accountability now or really hasn't been because not all hospitals need to follow the

standards every single day.

So, we know that the measure is the best way for there to be data collection on the process and so that is, in terms of usability, is the argument in the case that we made today.

DR. NISHIMI: Arjun.

MEMBER VENKATESH: I just feel a little nervous where we are on this usability space and the role of an NQF Steering Committee and the reason is this: technically, the measure is not in use.

It is not being used right now. There is no evidence of oh, in the presence of doing the screening, we had this quality improvement. It wouldn't meet the thresholds here. Where we are at is in a space where, hey, the Joint Commission has a -- we are climbing around this and CMS wants a measure like this for the Inpatient Quality Reporting Program. And so we can choose to endorse this measure and say oh, it has great usability potential because of the Inpatient Quality Reporting program but then

those programs will turn around and say look, this measure is NQF endorsed; it is a good measure. And so you can kind of end up in this cycle or circle of saying the potential for use or the ability for the measure to actually drive the development of a quality improvement agenda is how endorsement is being used.

The risk there, though, is that we then start endorsing measures or selecting measures on criteria different than what is originally set out. What is originally set out is does it have an evidence base right now? Does it have testing right now? Does it have feasibility and usability right now?

And so this is -- I feel like it is a real dangerous place to go to say oh, because there is so much interest in the use or because if we endorse the measure, the quality problem will get solved is a dangerous place to go with endorsement, I feel. In general, I feel like there is other mechanisms in the measure and policy space for that, the MAP, other things.

But I just want to bring that up because people complain on the other side. People say oh, why did this endorse measure of lower evidence or why is CMS using this measure in an IQR Program and CMS is going to say well, it was NQF-endorsed. And so we are in that circle.

MS. MUNTHALI: So, Arjun, you brought up a good point. As part of our criteria, we don't just ask for current use. We do ask for planned use. And we expect that the next time the measure is up for maintenance, there is specific use, it has been implemented in programs.

So, to the extent possible, if there is any way you can, perhaps, update your submission with what you have expressed to us and what is missing in the submission, you would definitely meet the criterion for that.

MR. VALLADARES: We can certainly do that.

MS. MCCAULEY: And just a follow-up.

I know you had that question about population health. Yes, that as we move forward, we have already been working with CMS with IMPACT and all the measures. And malnutrition is a huge issue with their pressure ulcers, their readmissions, and we want to make sure that we do move to post-acute care and that is rehabilitation, long-term care, hospitals, your home health, as well as the skilled nursing facilities.

And so this measure is also being looked at in that vein. And it is going to go with our population health dieticians, community nutritionists. It is going to move in that spectrum.

CO-CHAIR QASEEM: So, let's vote.

DR. NISHIMI: Yes.

MS. CRAWFORD: Voting is open on Measure 3087, usability and use. One high; two moderate; three low; four insufficient information.

Okay, we have zero high, eleven moderate; five low; zero insufficient

information. So, 69 percent moderate, 31 percent low.

DR. NISHIMI: Okay, we are ready to vote on overall suitability for endorsement. So this is the committee taking its entire discussion on the measure. Voting on Measure 3087 --

MS. CRAWFORD: Oh, we can't -- no, wait. I'm sorry, you had consensus not reached on evidence.

DR. NISHIMI: Oh, on evidence! It was so long ago. Sorry.

So, we can take a break for -- what is the time?

CO-CHAIR QASEEM: Oh, no, no, no. We are going to keep going.

DR. NISHIMI: Okay.

CO-CHAIR QASEEM: We are break but for this measure we can move on. Right?

DR. NISHIMI: Okay, yes.

CO-CHAIR McINERNY: Does that mean that we will revisit this after the open

comments?

DR. NISHIMI: Yes.

CO-CHAIR QASEEM: Okay.

DR. NISHIMI: Sorry. It really was a long time ago. I forgot about that one.

CO-CHAIR QASEEM: We can keep going. Is that okay?

MS. MUNTHALI: Yes, lunch isn't here yet. So, I would suggest we keep going.

CO-CHAIR QASEEM: So, they are not going to feed us until we finish this. So, 3088, Completion of Nutrition Assessment. Ron or anyone want to take this on?

If you -- you are going to introduce it but you will have to keep it within one minute.

DR. NISHIMI: Anything that is new.

CO-CHAIR QASEEM: Go for it. Your clock starts.

MS. MCCAULEY: So our next measure submitted for your review is fully specified for use with EHRs measuring the proportion of

nutrition assessments completed for patients at-risk of malnutrition identified by a completed malnutrition screening.

So, this is NQF 3088. Nutrition assessment is recommended for patients who are identified to be at-risk for malnutrition by screening. This measure focused on elderly patients, age 65 years and older who are specifically at higher risk for malnutrition due to more prevalent comorbidities such as COPD, dementia, orthopedic conditions, and some forms of cancer.

A 2014 study by Snider, et al of all the burden of malnutrition on elderly in the United States demonstrated the prevalence of malnutrition in the hospital is as high as 38.7 percent. The completion of a nutrition assessment using a recommended assessment tool, such as our nutrition-focused physical exam or a subjective global assessment provides the opportunity for a registered dietician to assess the patient for physical findings of

malnutrition.

This process allows for a malnutrition diagnoses and also informs the development of a nutrition care plan and includes the proper evidence-based intervention for the patient, based on their assessment needs and results.

CO-CHAIR QASEEM: Thank you. Ron or Cathy? Any comments? Any general comments?

Oh, is it --

DR. NISHIMI: Arjun.

CO-CHAIR QASEEM: Oh, it is Arjun. Sorry about that. Sorry.

DR. NISHIMI: Oh, yes, Arjun and Cathy, still.

MEMBER VENKATESH: This is just the next step in the measure in the sense that if the denominator obviously used if you screen positive and are identified as at-risk for malnutrition, then was a structured assessment done.

I think we discussed I think the structured assessment enough before. There was broad agreement in that kind of any of the tools

is probably a meaningful way to do this.

For the evidence, itself, I think that there is a -- I think generally this is either probably a moderate, based on the fact that there is a clinical practice guideline with Grade C evidence or it could probably also might fall in that bucket of insufficient with exception, meaning that there is a good international or national consensus statement that supports the activity without the evidence.

I mean that is probably the better place for this because there a systematic review that said that included studies that there was limited evidence within those. Part of that probably has to do with the things you guys spoke about, the quality of the evidence earlier on.

And so I think it is either moderate or insufficient. I think either way, from an evidence perspective it probably moves forward in evaluations.

CO-CHAIR QASEEM: Cathy, do you have any comments?

And so we are getting into the evidence piece. Is it okay if we continue with the evidence?

DR. NISHIMI: Evidence?

CO-CHAIR QASEEM: Yes. Any other comments?

So, I have got a few comments. Again, I reviewed this measure a while ago, so I am going to be trying to be looking at it again. So, you do talk about, of course, there is variation in screening. There is no argument. I absolutely agree with you.

There is the variation in treatment of malnutrition, it is presumed in the document throughout. I did not see any evidence that there is variation in treatment. I am talking about the evidence piece. Okay? I know what is happening out there. That is different. But if you are going to look into the evidence, that is one issue.

Because you cite the evidence for this measure, which is a 2011 guideline by American

Society of Parenteral and Enteral Nutrition, which states that nutrition support intervention is recommended for all patients identified by screening assessment at risk for malnutrition, malnourishment. This is rated as a Grade C recommendation.

Once I started digging into the evidence, and your point taken in terms of the issues with the observation and randomized trials, the guideline is based on three small randomized controlled trials and one nonrandomized cohort study with historical controls, and one nonrandomized cohort with some issues as well.

And although I think this measure starts getting better compared to the other one that we just talked about, my concern is that there was significant biases that were very evident in this clinical practice guidelines, as well as the systematic review that was done. The risk of bias was very, very high. So much so that evidence is being used to behind this but

the evidence itself is -- I would caution the evidence that was presented.

And then evidence is a little bit inconsistent that screening leads to referral for nutritional intervention but I am not going to talk too much about evidence. The bottom line is I had some concerns in terms of the guideline and the evidence that was presented as the basis for this recommendation.

DR. NISHIMI: Any other committee comments?

MR. LYNCH: If I could just get you to clarify. When you are talking about the concerns you have, do you have any concerns about the implementation of interventions?

CO-CHAIR QASEEM: The evidence that was presented as basis for nutritional assessment for this measure. Evidence that forms the basis for this measure has got a lot of red flags.

That is why I actually went into the evidence, so you guys have it and you can look it up in terms of the evidence that you talk about

it. I listed the studies that is used. That is why, just to make sure, I really went into each guideline and systematic review that was used as evidence as basis for this measure.

So, not only those studies were questionable, the risk of bias, the risk of bias of the guideline by itself was, at least to my knowledge and my experience, I would classify it as very high.

MEMBER STIEFEL: Can you elaborate on the risk of bias?

CO-CHAIR QASEEM: So risk of bias, essentially bottom line is the evidence that was presented, those individual four studies that is based on, they went above and beyond what those studies were saying in terms of the folks who were on the panel who develop these guidelines. I would caution a little bit that some bias is related to that.

Of course, it is always the case with many of the guidelines. We all know that. But I was getting a little I think some red flags that

just kept on going up. One or two I am okay but when in order of the things you need to develop a good clinical practice guidelines, when one, two, three, four, five occur each of them, without getting into too much detail, started going up, I started getting worried that if there is even enough evidence to support this measure.

DR. NISHIMI: Marcel and then -- okay.

MEMBER SALIVE: So, I think I am going to go out of order again. I think the next one I have -- and I think by going in this order, we are kind of losing the forest to the trees. The next one it talks about nutrition care plan and talks about an intervention.

So, I think when you talk about screening and getting a diagnosis, you can get very narrowly focused on that. But the big picture on screening is generally to do that screening, get the diagnosis, make the intervention, and improve health outcomes. And so maybe you are right. I don't know. To me, six studies with a low risk of bias is actually a

lot.

But I looked at the other, the next one and there are -- you know I guess this guideline looks very feeble but there is a Cochrane Review on nutritional supplements for people who are malnourished that is very strong, I think, and says there are interventions for these people that do improve health outcomes and has over 10,000 patients. And there is also a recent randomized trial that is not in there that has another 600 people.

So, to me, the benefits are important to consider there and you know I think it goes together. And so you know that measure -- I felt the evidence for that measure is sufficient, moderate, whatever you want to call it.

And so here we are kind of in the middle and we can discuss the evidence on this in the middle. But really, the big picture is where they get the diagnosis, they get some supplementation, something happens, and you improve health outcomes. So, reduced

complications, improved health outcomes. I saw that and I was fairly happy with that evidence.

CO-CHAIR QASEEM: So, just to respond and Arjun, I will come to you, I did look at some of the other studies and trials as well.

So, the evidence, to my knowledge -- again correct if I am wrong. I looked at it, of course, as quickly as I could because I had a lot of measures to review. Evidence is focused on UTIs, pressure ulcers, falls, fractures, acute respiratory tract infection in those elderly patients. That is the trial that you are referring to.

You are not looking at, again, keep it in mind you are looking at beyond that right now. So, that is what my --

MEMBER SALIVE: No, I think the Cochrane Review had 24 trials and it wasn't just those patients that you are talking about. So, there was broader -- there were multiple other.

CO-CHAIR QASEEM: Arjun?

MEMBER VENKATESH: Just so I have it

right, we are doing 88 right now, right? Okay, so for 88, I opened up the guideline, the 2011 one that is cited. There is one inaccuracy I think we should be aware of. This is -- what we are reading in the measure worksheet says that it was graded Grade C. In the actual guideline, the nutrition assessment is Grade E. What is Grade C is the intervention.

And so, just so we are clear, what is in the guideline specific to the measure, is Grade E, which is the lowest level grading.

CO-CHAIR QASEEM: Thanks, Arjun.

MS. MCCAULEY: Alison Steiber wanted to respond. Alison?

DR. STEIBER: Hello. I apologize. I'm in a little bit of a noisy space here but I just wanted to talk about the risk of bias issue that was brought up.

You know so when that study was done, there was not a real validated tool for risk of bias for nutrition studies. So, while I agree that certainly it is important to assess risk of

bias and certainly the grade methodology indicates that, I actually think that that is a bit of a limitation, as it relates to our studies and how we grade nutrition in systematic review.

And so I think we have to, instead, look at the more preponderance of information as it relates to this topic and the fact that there are quite a bit of studies that show improvements in outcomes with nutrition intervention, even if we don't have the grade level the way we would like it to.

And I think newer systematic reviews will do that, as we get better and better at assessing risk of bias in nutrition.

CO-CHAIR QASEEM: Thanks, Alison.

Any other comments on evidence? Anyone who is disagreeing with what we talked about, in terms of evidence?

DR. NISHIMI: So, I just want to clarify. The submission referred to part of the guideline and said Grade C. And Arjun has looked up the guideline and made reference to a

different part of it, referring to Grade E, which is substantially different from C, obviously.

So, what is the developer's position on what is the applicable guideline and, therefore, grade?

MR. VALLADARES: Thank you for that clarification. I think looking at the guideline or so myself, I think we did include the incorrect guideline recommendation. So, it should be nutrition assessment is suggested for all patients who were identified to be a nutrition risk by nutrition screening and that grade level is E.

DR. NISHIMI: Okay I just wanted to clarify it for everyone.

MR. VALLADARES: Thank you for that clarification.

CO-CHAIR QASEEM: Thanks so much, Arjun.

So, let's vote on evidence, please.

MEMBER SALIVE: Can I just say one last thing? Because I think that the other thing

to consider is just the risk of harm from this labeling. And that, I think, is extremely low in this instance also.

So, maybe the evidence is low but that is another consideration for screening is the risks of harm. And I don't see any evidence of that.

CO-CHAIR QASEEM: So but that was taken into account when they come out with the guideline recommendation as E. They look at benefits and harms, right? Except labeling, they might have looked at it.

Did they look at the labeling in that guideline, do you know? I mean you looked at all of the benefits and harms and then decided on what you are going to give it as a grade.

MEMBER SALIVE: It is not related.

CO-CHAIR QASEEM: Okay, so let's just vote.

MS. CRAWFORD: Voting on Measure 3088, evidence. One, high; two, moderate; three, low; four, insufficient.

One more vote.

MS. OGUNGBEMI: Mike, are you still with us?

MEMBER BAER: Yes, I did send one. I will send it again. Did you get it?

MS. CRAWFORD: Okay, so, zero high, eight moderate, five low, three insufficient. So, 50 percent moderate, 31 percent low, 19 percent insufficient.

So, it is consensus not reached. So, we will go on to discuss and vote on gap.

CO-CHAIR QASEEM: Any discussion on gap? Arjun, Cathy, anything?

MEMBER VENKATESH: So, I think they had some survey data that showed evidence of a gap. I don't think that there was testing data, necessarily, on the gap. But there was, I think the exact figure was 23 percent of the time out of a 1700-person survey was a structured assessment tool used.

So, it is a little tricky to apply because it is provider level not a patient level.

But I imagine that it feels right in terms of there being a gap in the use of a structured assessment tool.

CO-CHAIR QASEEM: Okay, seeing no -- well, we can probably vote on this one. No other comments? All right, let's vote.

MS. CRAWFORD: Voting on Measure 3088, performance gap. One, high; two, moderate; three, low; four, insufficient.

We have three high, eleven moderate, one low, zero insufficient. It is 20 percent high, 73 percent moderate, 7 percent low.

DR. NISHIMI: So, we should go on to discussing the validity and reliability testing.

MEMBER VENKATESH: So, they have got some EHR reliability testing done that shows that this can be specified in the EHR, that these data elements are captured, as well as related validity testing with chart abstraction.

I guess I am in a spot where -- and I don't know if this is out of order to ask the question but I feel like this measure could get

NQF endorsement if, on the first question, everybody totally understood like the opportunity to say that you can rate it as insufficient evidence but with exception because they are probably going to meet our standards with respect to reliability, validity and all these things down the line.

And so the real question -- the way the voting goes is when you get like moderate, low, and then some insufficients, you end up with yes, no consensus. But it is not a true like three-way branch point, like a low, medium and high is.

And so I am just wondering if we are kind of doing a disservice to this measure by not fully kind of giving it that insufficient with exception evaluation. Because then all these subsequent things are very different.

DR. NISHIMI: I mean the committee could, I think, at least decide not to continue discussing these elements and wait until the post-comment call when you settle the evidence

question and then continue down. That is an option.

MEMBER HILL: But we don't have the exception option?

DR. NISHIMI: No, you would reconsider starting at the top because you did not reach consensus on it. So, that is an option to just stop discussing this one, which is, I think what Arjun is suggesting.

MS. MUNTHALI: How many -- I would like to get a sense from the rest of the committee if you feel the same way that Arjun does and would like to stop or -- I mean if there is significant concern about your initial vote on evidence, maybe we revisit that. But I just want to get a sense from everyone on what their feelings are, their thoughts on this.

A show of hands. So, did you agree with what Arjun had just stated?

CO-CHAIR QASEEM: Yes, how many of you are like Arjun?

MEMBER SELLERS: I'm sorry to do this

but because I stepped out for a second, can you recap what you said quickly?

MEMBER VENKATESH: Sure, what I was just saying was that when we vote like medium, low, and insufficient, it creates this kind of false three-way thing, where really the first question should just be insufficient or not. And if we chose -- my guess is that some of the people who voted low might have actually chosen insufficient.

So, if those people who actually chosen first insufficient, we would then not would have voted to rate it high, medium, or low, we would have voted to rate insufficient with exception or not.

And I think this is actually the kind of thing that falls in that category, because there is not evidence. There is an expert consensus statement. There is not another rule measure out there. And we could have had a discussion about the risks. Is it on balance? Is this worth moving forward when there is an

expert consensus statement, in the absence of evidence.

But instead, we do this weird thing where we vote between low, medium and insufficient and so we ended up with this kind of smattering of votes between the three that doesn't allow us to go down that road at all.

And so I guess the question would be do you want to -- I guess I can make two proposals. Proposal one is we could vote on whether or not you would consider insufficient with exception. And if overall, there is some group consensus for that, maybe we can defer to the post-call. Or to just vote insufficient or not. And if a lot of people -- if there is consensus on insufficient, then you can always have the exception vote after.

MS. MUNTHALI: Yes, so Arjun, I just wanted to clear up something. You did follow the process. What we probably should have clarified is that if you did vote insufficient, then it takes down that path. But you have to have those

four options there.

But I don't know if we sensed that that is where the committee wanted to go generally, that you were thinking that it is not that the evidence was low, it is just that we didn't have sufficient evidence. And then, that would have given you the option of an insufficient with exception.

So, I guess the question could be maybe we just revote on that, knowing that like what we did yesterday to be consistent with the process, if you want the insufficient with exception pathway, you must have 61 percent of the committee voting at insufficient; not low, but insufficient.

DR. NISHIMI: Because we did have moderates and lows. And so it was all spread out.

So, let's see a show of hands if the committee wants to revote on evidence.

CO-CHAIR McINERNY: No. I mean I think we had a good vote on evidence and a good

discussion. So, you know I know we are in Washington, D.C. and certainly Congress is prone to revote, and revote, and revote but let's not follow their example.

DR. NISHIMI: Anyone else?

MS. MUNTHALI: We saw a couple of hands but don't see too many.

DR. NISHIMI: Okay. So, then consensus is not reached on evidence. We will come back, obviously we have to come back to revisit because we are not going to vote on overall suitability for endorsement.

The additional question that Arjun raised was whether this pathway, whether we should discuss now validity, reliability, usability and feasibility, or whether you want to hold that for a post-comment call. That was the second proposal he made.

So, let me ask for this. Show of hands to stop discussion.

MEMBER MOLINE: I would like to just --

DR. NISHIMI: Okay.

MEMBER MOLINE: I would just like to -- let's complete this thought process, rather than losing this train of thought for a post-call, if we can do this in a reasonable time frame.

DR. NISHIMI: Okay, I don't see any groundswell to suspend this measure. So, then let's continue with scientific acceptability of the measure properties. So, we are discussing reliability of the specifications, testing, and then the validity, which obviously goes to evidence.

CO-CHAIR QASEEM: Any more discussion on reliability and validity? I think we can do those two together, to a certain degree and then vote on two together. I mean both separately.

Any discussion on reliability and validity? None. Let's vote.

MS. CRAWFORD: Voting is open on Measure 3088, reliability. One, high; two, moderate; three, low; four, insufficient.

DR. NISHIMI: Actually, I'm sorry. The highest eligible rating here is moderate. Okay? So, the highest -- so, one, moderate; two, low; three, insufficient because they did data element level.

MS. CRAWFORD: You started voting? Okay.

CO-CHAIR QASEEM: Is that okay?

MS. CRAWFORD: That's fine, as long as we get 16 votes.

DR. NISHIMI: So, if you were voting on the other one, just re-press based on the one, two, three. It will re-record your new vote, so to speak.

MS. CRAWFORD: Okay, we have our 16. Okay, we have 13 -- oh, there is that. Well, I counted 16 -- they haven't voted yet?

Let's redo, just in case the two votes that were cast earlier we couched on this slide. So, if we could redo it one more time.

Okay, it's open. One, moderate; two, low; three, insufficient.

Okay, we have a total of 17 votes. Okay? All right, perfect. Fourteen moderate, three low, zero insufficient. So, 82 percent moderate, 18 percent low.

DR. NISHIMI: So, we will do validity. Again, the same thing. The high is not eligible.

MS. CRAWFORD: One is moderate, two is low, three, insufficient.

Okay, we have twelve moderate, three low, two insufficient. So, 71 percent moderate, 18 percent low, 12 percent insufficient.

CO-CHAIR QASEEM: So, we continue on feasibility?

DR. NISHIMI: Yes, feasibility.

CO-CHAIR QASEEM: Feasibility. Arjun, Cathy, anyone, actually.

MEMBER VENKATESH: No concerns. I think we have talked about this before. I think you have got the EHR data specified --

CO-CHAIR QASEEM: By the time you get here, you can just do it anyway.

MS. CRAWFORD: Voting is open for

feasibility. One, high; two, moderate; three, low; four insufficient.

We have one high, fifteen moderate, one, low. So, 6 percent high, 88 percent moderate, 6 percent low.

DR. NISHIMI: Okay, usability and use.

MS. CRAWFORD: Voting is open. Measure 3088, usability and use. One, high; two, moderate; three, low; four, insufficient information.

Okay, we have zero high, fourteen moderate, three low, zero insufficient information. So, zero percent high, 82 percent moderate, 18 percent low, zero percent insufficient information.

CO-CHAIR McINERNY: And so we revisit this, then, after the open comment period to try and resolve our consensus not reached on evidence. Correct?

MS. CRAWFORD: Yes.

CO-CHAIR McINERNY: Thank you.

CO-CHAIR QASEEM: Do you want to take

a lunch break or do you want to just ask how does everyone want to proceed?

MS. MUNTHALI: Well, I think we can do a 15-minute lunch break and that puts us back on schedule. So, we come back at 1:15.

CO-CHAIR QASEEM: A working lunch?

CO-CHAIR McINERNY: Working lunch.

MS. MUNTHALI: Working lunch.

CO-CHAIR QASEEM: But our measure developer folks, you will be able to stay?

MS. MUNTHALI: And we also neglected to ask if there are any public comments.

Operator, if you can open up the lines for the members of the public that are listening in that have a comment and anyone in the room who may have a comment on any of the measures we have reviewed.

OPERATOR: Okay, if you would like to make a comment, please press star and the number 1. You do have a public comment from Meredith Ponder.

MS. MUNTHALI: Hello.

MS. PONDER: Hi. Can you hear me?

MS. MUNTHALI: Yes, we can. Please go ahead.

MS. PONDER: Okay, great. Hi, I am commenting on behalf of Defeat Malnutrition Today, which is a coalition of over 40 organizations and stakeholders and we share the goals of achieving the recognition of malnutrition as a vital sign of older adult health and we are working to achieve a greater focus on malnutrition screening and intervention.

And we would like to say that older adults are at high risk of becoming malnourished and under nourished due to chronic illness, disease, injury, or social determinants, which makes it harder for them to recover from surgery and illness, makes it more difficult for their wounds to heal, increases their risk for infections and falls, and decreases their strength that they need to take care of themselves. And their health costs can be 300 percent greater than those who are not

malnourished on entry to the healthcare system.

And we support NQF endorsement for these four malnutrition quality measures, as it is critical to ensure that malnutrition is identified, treated, and that patient nutritional status is documented as a diagnosis in the patient's medical record to ensure prompt nutrition intervention and continuity of care for older adults upon discharge to home or post-acute care settings.

MS. MUNTHALI: Thank you.

CO-CHAIR McINERNY: Thank you for your comment.

MS. PONDER: Thank you.

DR. NISHIMI: Okay, I think we are ready to take a 15-minute break to go get the lunch and bring it back and we will have -- the rest of it will be a working lunch.

CO-CHAIR McINERNY: We need to keep well-nourished.

(Whereupon, the above-entitled matter went off the record at 12:59 p.m. and resumed at

1:20 p.m.)

MS. MUNTHALI: Okay, we're going to get started. And what we are doing right now is Yetunde is going around to our new committee members, Matt and Steve and Barry-Lewis, in absentia, Ann DeBiasi, we are going to pick the next terms for the committee. And when she is done with that, we will get started with review.

MS. OGUNGBEMI: If you could say your name and announce your choice, please. Thank you.

Oh, I had the newest committee members pick terms. So two or three years. Yes. Choice, yes. No switching.

MEMBER STIEFEL: And what are we supposed to do?

CO-CHAIR McINERNY: Read the number.

MEMBER STIEFEL: Three.

DR. NISHIMI: Matt Stiefel, three. And then Steve.

MEMBER TEUTSCH: Three.

(Simultaneous speaking.)

MS. OGUNGBEMI: There were an even

number of choices.

(Simultaneous speaking.)

MS. OGUNGBEMI: I picked for our other new committee member Anne De Biasi and she got two years. Thank you.

MEMBER MOLINE: Can I ask how do we know when our terms end? Because I can't remember when it was that we pulled the short end or long end, however we want to call it.

When we decide if it's feasible or usable?

MS. MUNTHALI: What we'll probably do is follow up with you after this meeting. We're tracking your terms and we'll let you know where you are and which one you picked.

I can't remember which you picked, but you're good for now.

DR. NISHIMI: Okay, we're going to proceed with our next measure if the developers can very briefly introduce their measure and then the committee will start discussing them.

If the comments are similar to what

you made, or exactly what you made for the other ones if you could be very brief about it so that hopefully we've discussed most of the issues with these measures and we'll be able to move through quickly.

MS. MCCAULEY: Okay, thank you. So this measure is 3089 and it is an eMeasure hybrid, partially specified for use with EHRs and partially requiring chart-abstracted measure data.

Again, this is for patients age 65 and older.

The nutrition care plan contains the registered dietitian's recommended approach for intervening on the patient's malnutrition.

The findings from the nutrition assessment serve as the basis for determining the appropriate way to address the patient's condition.

The measure is not oriented toward a specific intervention, but rather a process whereby the assessments and interventions are

organized for implementation through a recommended plan of care.

As of late 2015 the majority of states have specific statutory or regulatory impediments that exist and preclude registered dietitian nutritionists from taking full advantage of ordering writing privileges.

And the CMS expanded that role for dietitians, but state by state we have to then make sure that that's implemented. And so that has hindered in some of the automatic interventions and ordering that need to be done.

The 2014 study of malnutrition care practices in the United States showed only one quarter of the surveyed clinicians in the United States reported whether nutrition assessment informed the determination of a malnutrition diagnosis.

The results of the study represented a downstream gap in care for determining the appropriate nutrition care plans as the findings of a nutrition assessment completed by a

registered dietitian inform such a plan of nutrition intervention for malnourished patients. Thank you.

CO-CHAIR QASEEM: Okay. Comments on this measure.

MEMBER SALIVE: So, I said earlier what I said, but I'll repeat that it was very good, I thought, that they had not just the guideline, but a Cochrane Review which had many trials supporting the supplement.

The nutritional supplement to these malnourished people does improve several different health outcomes.

So, I thought this evidence spanned the full workflow that they outlined, and that it was strong.

CO-CHAIR QASEEM: Other comments? So, I see what you're saying, Marcel.

When I looked at this measure quickly, at least the evidence that's presented by the measure developers, and please correct me if I'm wrong. And Alison, hopefully you're still on the

phone as well.

The performance gap, I'm going to go through all of the comments together because we're going to probably move quickly through it.

It's from a paper that was published in 2008. So, I am not really sure why the measure developers did not look at any newer data when they were talking about the performance gap.

Again, as I said I'm going to cover all of my comments together so we can keep the process moving.

And the same paper also showed that the patients who had this assessment were more likely to get additional feeding and vitamins, but they did not report any difference in outcomes.

It includes a level C recommendation and that's based on a systematic review that was again in the seven or eight years ago for which the data was collected between 1986 and 2005. So you're looking at almost 11 years old the latest data that was used in this one. Just a few

general comments.

So, I understand what you're saying, and since we're not discussing Cochrane, because I don't want to waste committee's time because I know that evidence as well. I've looked at Cochrane Review.

But based on what you have presented in the -- I'm just curious why are we going that far back for gap data as well as for evidence.

Even the gap is from 2008.

MR. VALLADARES: So, we again used the Cochrane data to support our evidence submission, our evidence attachment.

And unfortunately the most recent broader guideline that is supported through some of the largest professional societies is that ASPEN systematic guideline.

And then the only other recent evidence is not systematic. It's the consensus statement. So, that's basically the most recent data that there is available at a systematic level.

MEMBER SALIVE: There was a 2016 trial which I mentioned earlier also. So I think there's some more recent evidence.

I think nutrition trials span the whole spectrum of duration. There's some older ones and newer ones.

I agree though about the gap, that that evidence was scant. But I believe there is a gap.

CO-CHAIR QASEEM: Other comments? Okay, how do you want to go about doing this? Shall we just go through the comments quickly about it in general and just vote on it, or just go through each step?

DR. NISHIMI: Yes, I think you just need to move to -- so, evidence. If there are any other committee members that want to address evidence or gap. Otherwise we'll vote on both.

CO-CHAIR QASEEM: Let's vote.

MS. CRAWFORD: Voting is open on measure 3089. One - high, two - moderate, three - low, four - insufficient.

We have 1 high, 14 moderate, zero low, 1 insufficient. Six percent high, 88 percent moderate, zero percent low, 6 percent insufficient.

DR. NISHIMI: Performance gap. Ready to vote.

MS. CRAWFORD: Voting is open on measure 3089, performance gap. One - high, two - moderate, three - low, four - insufficient.

We have 1 high, 11 moderate, 1 low, 2 insufficient. That's 7 percent high, 73 percent moderate, 7 percent low, 13 percent insufficient.

DR. NISHIMI: Scientific acceptability, reliability and validity testing. They did data element so it will ultimately be eligible only for the moderate rating.

But was there anything different about this measure that the committee members who reviewed it want to make a comment on?

MEMBER MINNICH: I would just add that, again, it was tested across three different EMR systems and two different sites.

DR. NISHIMI: Did you want to expand on whether you thought that met the reliability or validity --

MEMBER MINNICH: I believe that it did, yes.

DR. NISHIMI: Okay. Any other comments on validity and reliability? Arjun, Amir.

MEMBER VENKATESH: I think they did basically data element reliability that has adequate agreement. The kappas are lower probably because the samples are smaller and so I would give it moderate.

DR. NISHIMI: Okay.

MS. CRAWFORD: Voting is open on measure 3089 reliability. One - moderate, two - low, three - insufficient.

We have 10 moderate, 5 low, zero insufficient. Sixty-seven percent moderate, 33 percent low, zero percent insufficient.

CO-CHAIR QASEEM: Any comments?

DR. NISHIMI: That was just

reliability.

CO-CHAIR QASEEM: Oh, sorry. I'm ahead.

DR. NISHIMI: Any other comments on validity? Arjun?

MEMBER VENKATESH: What are the standards for this for an eMeasure that's kind of in the works? Because they only have two sites. And so I don't know for sure -- the main one that I always care about in validity is can you make meaningful inferences about quality between two different facilities based on the score.

They have testing data from two sites that is different. That doesn't mean that you can actually detect meaningful differences, but that may be an undue expectation for an eMeasure that doesn't have a bunch of testing data that would do it.

And so --

DR. NISHIMI: The eMeasure requirement was at the two sites. But the committee could decide that, you know, you want a higher

threshold.

But the minimum for purposes of eMeasure feasibility is really what it goes to, not the question of whether there's a meaningful difference.

MEMBER VENKATESH: Gotcha.

MEMBER HILL: So can we clarify whether the hospice exclusion is included or excluded of the -- with regards to the final specifications.

MR. VALLADARES: Great, thank you for that question.

So, in this particular measure following along sort of that logic that I shared earlier there was a significant difference in the results for both the data element and the performance score level when the patients that met exclusion criteria were excluded.

So, in this measure the specifications do exclude all of those patients that we -- so it was the hospice, discharge against medical advice, and also length of stay under 24 hours.

DR. NISHIMI: Any other questions or comments? Okay, ready to vote on validity. Moderate, low and insufficient are your options.

MS. CRAWFORD: We have nine moderate, seven low, zero insufficient. Fifty-six percent moderate, 44 percent low.

DR. NISHIMI: So, consensus is not reached on the validity and we'll continue to discuss feasibility and usability and use, but we'll hold the final vote until after comments are received on the post-comment call where you will review those comments on then vote on the post-comment call.

CO-CHAIR QASEEM: Any comments on feasibility?

DR. NISHIMI: I just want to clarify for the committee this is different because it involved chart review as well as EMR.

CO-CHAIR QASEEM: That's okay. I mean, it's just -- you know what I'm going to say about EHR review.

MS. CRAWFORD: Okay, voting is open

for measure 3089 feasibility. One - high, two - moderate, three - low, four - insufficient.

We have five high, nine moderate, two low, zero insufficient. Thirty-one percent high, 56 percent moderate, 13 percent low, zero percent insufficient.

DR. NISHIMI: Okay, usability and use. Any committee comments? Okay, we can vote on usability and use.

MS. CRAWFORD: One - high, two - moderate, three - low, four - insufficient information.

We have 2 high, 11 moderate, 2 low, 1 insufficient information. Thirteen percent high, 69 percent moderate, 13 percent low, 6 percent insufficient information.

DR. NISHIMI: And we won't vote on this measure until the post-comment call. So thank you very much, developers. You had to slog through a long session. We appreciate it.

We're going to move on to 3039, preventive care and screening, BMI screening and

follow-up.

Measure developers here? Is the measure developer on the phone? Oh, they're making their way up. Okay.

If you could introduce your measure, two to three minutes, I appreciate it. And then the committee will discuss and if they have questions we'll engage.

MS. SOMPLASKY: My name's Anita Somplasky from Quality Insights and with me is KeriAnn Wells from Mathematica Policy Research.

On the phone we should have Dr. Dan Green from CMS who will also be answering questions.

We are pleased to introduce NQF 0421: Preventive Care and Screening: Body Mass Index Screening and Follow-up Plan for consideration for NQF re-endorsement.

We will discuss two versions of the measure - NQF 2828 is the electronic clinical quality measure, and NQF 3039 is the claims and registry version of this measure.

This measure was first implemented in a CMS quality program to promote healthy weight by screening patients for BMI scores and identifying patients appropriate for an intervention. That is, those outside of normal parameters.

The measure was first implemented in the Physician Quality Reporting System (PQRS) in 2008 and it was added to the Electronic Health Record Incentive Program, commonly referred to as Meaningful Use in 2010.

The intent of this process measure is that all eligible professionals document a patient's BMI during an encounter, or in the six months before the visit.

When a patient's BMI is outside of normal parameters the measure requires that eligible professionals document a follow-up plan such as exercise, nutritional counseling, or a referral to a specialist to help the patient achieve a healthy weight.

This measure focuses on adults and

includes all visits during the 12-month reporting period.

As stated in the 2013 AHA/ACC/TOS Guideline for the Management of Overweight and Obesity in Adults the biomedical, psychosocial and economic consequences of obesity have substantial implications for the health and well-being of the United States population.

More than one-third at 34.9 percent of adults in the United States are obese. Obesity among adults younger than 65 has been shown to reduce life expectancy and increase medical costs.

Weight loss has been shown to decrease blood pressure, reduce triglycerides and decrease blood glucose levels and hemoglobin A1c, all of which may slow the progression of type 2 diabetes and cardiovascular disease.

Unfortunately fewer than 50 percent of obese adults in 2010 received advice to exercise or perform physical activity.

On the other end of the spectrum the

2007 to 2010 National Health and Nutrition Examination Survey indicated that an estimated 1.7 percent of adults in the U.S. ages 20 and older are considered underweight.

Elderly patients with unintentional weight loss are at higher risk for infection, depression and death, but these concerns can be alleviated through counseling and monitoring.

This measure reflects an important aspect of care that clinicians do not regularly provide.

The average 2014 PQRS performance rate was 61 percent with fewer than 20 percent of eligible professionals reporting.

This measure has the potential to alert eligible professionals to the importance of identifying populations at risk and treating patients using evidence-based guidelines, maximizing population health and reducing healthcare costs.

We thank you for your consideration and look forward to your questions and

discussion.

DR. NISHIMI: Before we go to the committee's discussion I did have a question for the developer.

This was previously endorsed as 0043 and it's been resubmitted now and is 3039.

Did the specifications change and that's why you resubmitted it instead of referring to it as a straight maintenance measure?

MS. SOMPLASKY: We were instructed now that it is an eCQM that we needed to have two new numbers issued which they were and we submitted it that way.

DR. NISHIMI: Okay, so then the evidence and testing data and all the other sort of if you will old maintenance still apply to this.

MS. SOMPLASKY: Yes, ma'am.

DR. NISHIMI: Okay, thank you.

MEMBER SPANGLER: Robyn, I have a question about that too. Did you mean 0421? Is

that right? Okay, so they're the exact same measure?

DR. NISHIMI: When I'm not wearing my glasses I can see neither far nor close apparently.

(Laughter)

DR. NISHIMI: But yes, thank you, Jason. So this really should be thought of as a maintenance measure.

CO-CHAIR QASEEM: Comments? Cathy or Matt or it doesn't matter. Go for it. Cathy.

MEMBER HILL: I was just wondering if there is any data especially since this is a maintenance measure around the variations in ethnicity in terms of the ranges of BMI that are considered obese.

MS. WELLS: So, we did test the measure in racial categories and we found some differences there. But there was no change to the measure. There was no risk stratification or adjustment on that basis.

CO-CHAIR QASEEM: Matt.

MEMBER STIEFEL: Just a comment. It's good to see a measure that includes screening, documentation and follow-up unlike many of the previous measures that have been so parsed out that they've been very difficult to evaluate. So this shows that it can be done and I appreciate it.

MEMBER SALIVE: So, I was one of the reviewers and I think it's strong. The evidence hasn't changed, but it's still endorsed by the U.S. Preventive Services Task Force.

They presented lots of evidence about the gap.

It seems quite feasible and I think although some people saw the 19 percent in PQRS as low, I believe it's because they can choose their measures.

And so it's actually in the top five of PQRS measures amongst ones reported. So, nothing is much higher than 19 percent.

So, hopefully this can get out and be more widespread, but I think they presented a

strong case for meeting all the criteria.

DR. NISHIMI: Tom.

CO-CHAIR McINERNY: Yes, I'm unclear about a couple of things.

Because I don't see adults, I assume most adults, whenever they come into the office for any reason they get weighed. Is that pretty much correct? Is that what internists do? So they get weighed.

And you can assume that their height is constant pretty much.

CO-CHAIR QASEEM: They check height as well.

CO-CHAIR McINERNY: But probably not as often, right? I mean, I've lost an inch over the years, you know, my intervertebral disc collapse.

But so you can -- if you know the height which you say is constant, and you have the weight then the BMI can be easily calculated.

Now, that's a little different in pediatrics. We don't do a height and weight at

every visit. We only do them at our health maintenance visits.

You know, if a kid comes in with an earache we don't do a height and weight so we don't do a BMI.

But I guess for adults that's not a big problem because you do it for every visit.

The other question I have is you say that there's a plan for follow-up. What do you count as a good plan for follow-up and how do you determine that meets the requirements for the measure?

MS. SOMPLASKY: We have not been prescriptive about what follow-up entails because this measure is meant for all eligible providers.

So, it's more than just physicians, it includes social workers, psychologists, nutritionists, physical therapists.

So, because of that we did not want to be prescriptive and say that it had to include certain testing.

So, a referral back to the primary

care physician is sufficient. If you are a primary care physician, or one of the specialists seeing them, following -- either recommending a weight management program or reassessment, something in that chart that shows that a follow-up, some sort of follow-up has been documented for BMI outside the parameters.

CO-CHAIR McINERNY: You're confusing me a little bit because you're saying social workers, psychologists. Did you say physical therapists?

MS. SOMPLASKY: Yes, sir.

CO-CHAIR McINERNY: They don't weigh patients on a regular basis.

MS. SOMPLASKY: For this measure they do.

CO-CHAIR McINERNY: What's that?

MS. SOMPLASKY: In order to report this measure which they do they are required to weigh and measure.

CO-CHAIR McINERNY: Okay. So the follow-up can be pretty open-ended like you're a

little overweight, or you're a lot overweight, whatever you want to say. Eat better, exercise more, come back in six months. I'll weigh you again. And that's considered sufficient.

MS. SOMPLASKY: Yes, sir.

CO-CHAIR McINERNY: Well, we know that doesn't work.

(Laughter)

CO-CHAIR McINERNY: I mean, come on, let's be honest about it. They come back in six months and you're lucky if they're the same weight. Often they're another pound or two heavier.

So, I'm a little concerned that that -- we need to make a better plan to help patients who are overweight. And just come back in six months I don't think really cuts it. I'm sorry.

DR. GREEN: Excuse me, can I jump in there for a minute? This is Dan Green from CMS.

So, I would agree with you. Ideally we would prefer to have a more robust plan for folks to follow.

But I think part of the intent of the measure is to get physicians and other caregivers engaged in confronting patients that are either dramatically over- or underweight.

So, just the fact that someone addresses it rather than they put them on the scale and they make no mention of it is better even if the patient doesn't do exactly what's recommended than ignoring it altogether.

So, I agree that it could be more robust, but I would also suggest that something is better than nothing.

MEMBER HILL: Thank you. So, just a follow-up to my question on diversity because I am aware of data that is out there, especially on the Asian population that we can't measure them the same way and identify obesity the same way as we have done in the past.

You mentioned that you did find some differences and that they were not included.

Do you have some plan to update your evidence on this measure so that our growing

Asian population has an accurate measure of whether they're obese or not?

MS. SOMPLASKY: We didn't find statistical differences. We did -- for what it was we did have around ethnicity.

And we are aware of the differences for the Asian population that for them they're going to have a much lower threshold than other ethnicities will.

But we weren't seeing any statistical difference in what we got in our results.

MEMBER CARRILLO: In our ACO we have a lot of experience with this measure. It's one of the required ACO measures.

And it's effective. I can say that particularly the follow-up piece. Because the BMI is pretty much baked in and it's part of many usual processes in doctor's offices.

And the height you catch once and assume that that more or less stays.

But the follow-up which in our case is prescribed as return visit, as counseling, as a

weight management program and a couple of other options is part of a dropdown.

So, physicians have to actually address in that dropdown what the follow-up is. So we find that we lose points in the follow-up piece.

So in terms of our education with doctors we're basically focusing them on you don't just do a BMI, you do something about it.

So, we have found that to be a challenging measure, but also a very effective measure for the same reason.

MEMBER MOLINE: I'm fixated on the social worker thing. I just, I mean I reviewed these measures and I didn't see anything that would make me remotely think that a psychologist or a social worker was being held to this measure.

And I've reviewed a ton of medical records in my career and can never remember seeing a social worker or a psychologist be expected or do these measures unless perhaps they

were a bariatric social worker or psychologist.

But I don't have experience with that.

So can you just clarify very quickly?

MS. SOMPLASKY: Sure. In the PQRS reporting program eligible providers in order to not receive a payment adjustment you have to report a set number of measures.

This measure when it was originally developed back in 2007 was developed with keeping all eligible professionals that report for the PQRS reporting program in mind. And that does include social workers, psychologists, physical therapists, occupational therapists.

They are actually part of our expert work group that we have for this measure so that we were not setting the bar too high for them.

But because we have to keep in mind all of those -- they call them the eligible professionals, but all of those specialties that are eligible to report for PQRS, they are included for this measure as well.

MEMBER BIALEK: The measure applies to

all providers. And whether they're privately -- whether the individual patient is privately insured, uninsured, et cetera.

So, the data that you have about reporting comes from a selective group. I didn't see anything about the ability for this standard to actually -- for this measure to actually be reported by others who aren't part of either the Medicare system or CMS type of system.

MS. SOMPLASKY: So, the Medicaid Meaningful Use program would include this measure as well. That's reported to the states so we did not have that data because that's reported to each state individually and they handle the Medicaid reporting program.

It is not -- there are BMI measures that are used in HEDIS and are used by other health plans, but they don't have the same measure we do with that follow-up plan, or the follow-up requirement included with theirs.

MEMBER BIALEK: But you're proposing this measure would apply though to everybody, not

just Medicare and Medicaid. Correct?

MS. SOMPLASKY: When I say all eligible providers for that PQRS reporting system.

MEMBER BIALEK: Okay. So the way I read the measure it doesn't specify that this is limited just to -- so is it just limited to the populations you mentioned, Medicare, Medicaid?

MS. SOMPLASKY: Anybody could use this measure if they so desired. We just haven't seen anybody outside of Medicare, Medicaid.

There's an NQF measure that actually used this as the basis, but they made it very disease-specific.

MEMBER VENKATESH: So, Ron, I think actually everybody is included. The measure denominator is all patients age 18 and older. So, everybody's in.

It's a pretty basic I think straightforward screening measure. It's hard to measure anything else really in this space.

There's a USPTF B-level recommendation

that they would meet moderate for evidence. The measure is being used.

Nothing really hangs me up about this measure. I think it's a no-brainer we should move forward.

I think one of the things just to remember about these measures is that the whole PQRS program that's proof of use of this measure, somebody's score on that is their best case score.

And so if there's still evidence according to this that there's still a gap in somebody's best case score.

And so chances are the actual scores are even worse. It's a program people can participate in by choice. It's not mandated.

But it's important that for those programs individual clinicians have measures available to them, and this is one that hits on a big health topic.

And so I kind of would say we should -- it's already endorsed. We should move

forward. They're trying to make an EHR version of this measure to make -- reduce the burden of data collection and kind of advance it. It makes sense to me.

MEMBER CARRILLO: Well, as was mentioned this is a Meaningful Use tool measure. So it's really anyone -- there's a lot of adoption of Meaningful Use by practices everywhere.

And whoever has access to the electronic medical record, be it social workers or physicians, nurses, whatever, will be actually having access and acting on this measure.

CO-CHAIR QASEEM: A couple of comments that I want to make after hearing all of what you said.

I think we need to be a little bit careful with what the task force recommends versus what the evidence -- what this measure is about.

The task force is very specifically screen all adults, but the follow-up plan is in

the obese population, not overweight. And there was a lot of discussion in the task force meeting about whether we include the overweight population or not.

That's one issue. And the evidence is actually a little bit not clear about it.

And the second issue is that you need that multi-component extensive follow-up plan. Going back to what I think Tom was bringing up that's important as well.

Just having a follow-up plan has not shown to have any impact on patient population.

And the third point I want to raise is the time issue. I was trying to pull the number and my dashboard was crashing on me, the NQF one. But I think it said six months if I remember correctly.

And there is -- most of the folks, expert -- there is no evidence. They said once every two years might be a better option to go, or at least once every year.

The reason being is that sustained

weight loss is important. Just again going back to what Tom brought up, I think it's an incredibly important issue, that every six months you're doing it is not going to really show or do anything in terms of improving health outcome.

And again, it's a very important issue. I'm not saying that this is not an important issue.

I think everyone in this room is going to agree obesity is an important issue and we don't just figure it out.

And some of this conversation, going back to 2014. I honestly don't remember which NQF committee I was on, but this happened at that point as well and CMS was asked to address some of these concerns.

And what we were told was next time around when the measure comes for endorsement purposes some of them will be taken into account, including issues such as BMI alone does not help.

A lot of new evidence that's out there in terms of waist circumference and waist-to-hip

ratio that has shown more direct correlation than just BMI itself.

And at that point in time the conversation was that if CMS is going to bring it back for discussion and when next time around the measure will be updated we're going to take that into account.

But I haven't seen any change in this measure. It's exactly what it was we discussed last time around.

So, can you address some of these?

MS. SOMPLASKY: In terms of changing for the follow-up plan we have not made any significant changes to that.

And one of the things that we'll talk about is that documentation in the EHR has proven to be very difficult to have that in a structured field and to be able to show that.

So, we know that there are challenges with that.

We did make recommendations for 2017 to have one population and not have the

stratification.

But because we were testing on 2014 data that we had our guidance from NQF was to use the measure that was consistent with the data that we have.

But we have not made significant changes to the measure.

CO-CHAIR QASEEM: What I'm saying essentially is it's not even consistent with the task force recommendation right now anymore.

The task force revised its recommendation and it does not include everyone like the way you have it in this measure. It doesn't match up with that.

I can read the task force recommendation right now if you guys want.

It says the task force recommends screening all adults, but the clinician should offer to refer patients with BMI greater than 30 or higher for intensive multi-component behavioral intervention.

If you start going into the clinical

consideration and implementation of it, it talks about exactly the issues that have been brought up by this committee at this point.

And this recommendation actually came out -- it's been awhile. So, it could have been incorporated by now.

MEMBER SALIVE: Any screening recommendation like that is a moving target in some ways. And how you translate it into a quality measure can be argued.

That's what I was trying to say this morning. And I think this is what brought it to me because I think there is some wiggle room here on this measure.

So, they are also addressing -- the underweight is addressed in this measure and the obese both.

But also there's an interpretation role of the physician. There are people, the muscular people who have a high BMI who don't need anything.

And you have to sift through that

also, and that's still a plan. The plan is do nothing, actually. They're muscular.

And the timing, I mean I think the USPSTF recommendations always have these subtle gradations that don't translate well into quality measures.

So, you know, I think it's a trade-off between -- they've tweaked the measure basically into two forms, the registry and claims one that we're talking about now, and the eMeasure.

I think it's somewhat popular from the evidence. Is it perfect? I doubt it's perfect, but I doubt we'll ever have a perfect one.

CO-CHAIR QASEEM: And I don't know if we'll be able to debate that today or not. That's a very interesting perspective, Marcel.

I agree with you. I'm not disagreeing with you. But what I'm saying is just because something is good clinical care do we have to have a performance measure for it? Does every single thing we do in our clinical practice needs to translate into a clinical performance measure?

That's one thing.

And I think I would have been more willing to go take that route until performance measures became a high-stakes game under MACRA by CMS.

If you're going to be measuring physicians and now you're going to be impacting their -- essentially it's pay-for-performance now then you need to come up with performance measures that are at least based on evidence, if you're going to go with evidence.

And we need to apply the evidence uniformly. What I heard today was some of the measures that we were approving we were saying oh, it's a task force recommendation. Let's go for it.

And then if that's what the standard is then let's stick with that standard. If that's not what the standard is let's apply -- let's not pick and choose standards based on which measure we are reviewing, or where is it coming from.

I want consistency across the measures, but that's what I'm struggling with.

There are many good performance measures that are out there that are good evidence base.

And remember that the guideline in performance measures -- it's all of your clinicians.

The guidelines, I can apply the guideline and make a judgment call. On performance measures it's not the case. Performance measure is saying you need to do this in every patient.

There's a huge difference between guideline and performance measure. And I think guidelines give you some flexibility. Performance measures don't.

And I'm not saying I'm disagreeing. Obesity is a huge issue. It's an important topic area, I'm not going to disagree.

Absolutely it's a good clinical practice as well. I'm not going to disagree with

that.

What I'm saying is that at least it should show what the current evidence is showing, and what -- and what the governmental agency is saying. Either we agree with task force or we don't agree with task force.

DR. GREEN: This is Dan Green again. I think you need to be a little bit -- have some blinders on in terms of the program and the higher stakes of quality measures.

To your point I don't think anybody can argue with the importance of assessing one state, particularly given the epidemic of obesity in our country.

Now again, I know the measure talks about underweight individuals as well, but let's focus on the overweight for just a minute.

The measure certainly may have room for improvement, but I believe there's a measure that's been NQF-endorsed that just assesses BMI.

So you know, that's like asking a cigarette smoker if they smoke and documenting

it. Well, that's great, but if you're never going to counsel them, or even make any effort to try to get them to get their smoking under control, or in this case their weight under control you've not really done anything.

It doesn't take a task force and it doesn't take a rocket scientist to figure out if somebody's overweight they should be at least advised and have some effort made to have them get their weight under control.

So, I'd hate to throw the baby out with the bath water so to speak as it relates to this measure.

It's an optional measurement -- you mentioned our programs. It's an optional measure for folks to report in our PQRS program and it's proposed to be optional under the new MACRA legislation.

So not making anybody do this measure. But I think it's an important measure that we continue.

CO-CHAIR QASEEM: And I absolutely

agree it's an important measure. I'm not arguing that.

I'm just pointing out there are a lot of things that could have been updated with this measure including as I said the follow-up plan alone. There is enough evidence out there that shows just follow-up plan does not change.

So if CMS really wanted to improve the patient outcomes we need to really work on what is going to change. Just follow-up plan has never shown.

And even in terms of the intensive therapies that you're talking about it takes 12 to 18 months to show any change.

And I'm not even talking about statistical change. And I can send you a lot of good evidence on this one if you look at the literature on this.

So, all I'm saying is this measure is saying six months, you need to do something every six months.

You're not talking about -- you just

have it as follow-up plan. It does not get into intensive -- or give some guidance which is going to make a difference.

Obesity is an important issue and let's do it the right way. Just having it as a checkbox as a measure that yes, we have a measure on obesity is not going to change the attention out there, at least in my opinion.

DR. GREEN: I agree a more robust plan would be beneficial, but we don't want to sit there and tell physicians and other eligible professionals how they should be providing care.

It's a guideline. Are you asking your patients and assessing your patients' weight.

And if you are and they fall out of the norms are you at least guiding the patient as to where they can get help, or how they can get help.

CO-CHAIR QASEEM: This is not a guideline though, I disagree. You have it as a performance measure. You used the term guideline.

If it's a guideline I'm perfectly fine with it.

DR. GREEN: I said are you guiding. I didn't say guideline. I said if you fall out -- yes, I did. I said if you fall out of the BMI norms are you sending the patient for -- giving them guidance to get help to try to get their weight under control.

CO-CHAIR QASEEM: All right. So, the last two comments and then we'll vote. Marcel and Emilio.

MEMBER CARRILLO: Just to say that perfection is the enemy of the good.

And if we don't have providers, physicians paying attention to people being out of range, up or down, the best laid plan for improvement is not going to do anything.

So, to get docs focused on the BMI which you would think they would. Our experience in our ACO is that they don't.

So I just think that it's a good first step.

DR. NISHIMI: Ready?

CO-CHAIR QASEEM: What are we voting on?

DR. NISHIMI: Evidence.

MS. CRAWFORD: Voting is open for measure 3039 on evidence. One - high, two - moderate, three - low, four - insufficient.

We have 2 high, 10 moderate, 3 low, 1 insufficient. Thirteen percent high, 63 percent moderate, 19 percent low, 6 percent insufficient.

DR. NISHIMI: Okay, the next issue is performance gap. We kind of did this discussion already so let's vote on gap.

MS. CRAWFORD: One - high, two - moderate, three - low, four - insufficient.

We have eight high, eight moderate. Fifty percent high, 50 percent moderate.

DR. NISHIMI: Okay. It passes gap. We can go onto scientific acceptability.

They did empirical testing for reliability at the score level so it is eligible for high.

It was a signal-to-noise ratio. That's typical of reliability testing. And the reliability statistic was 0.75. So that's generally considered good.

MS. CRAWFORD: Voting is open for measure 3039 reliability. One - high, two - moderate, three - low, four - insufficient.

We have 10 high, 6 moderate, zero low, zero insufficient. Sixty-three percent high, 38 percent moderate.

DR. NISHIMI: So it passes the reliability criterion.

For the validity testing it is data element level, manual abstraction against an EHR extract. So, as data element validity testing it is eligible for a moderate rating.

The agreement on the numerator was 90.16 percent. The kappa was 0.8. And for the denominator it was 99 percent and the kappa was low, 0.4, but the developer attributes that to the low number of exclusions in the extracted data.

So, if the committee has any discussion on that. Ready to vote?

MS. CRAWFORD: Voting is open, measure 3039 on validity. One - moderate, two - low, three - insufficient.

We have 12 moderate, 4 low, zero insufficient. Seventy-five percent moderate, 25 percent low.

DR. NISHIMI: Okay, so the measure passes validity.

Onto feasibility. We've already discussed the feasibility to some extent. Are there any additional comments?

Okay, let's vote on feasibility.

MS. CRAWFORD: Vote on measure 3039 on feasibility. One - high, two - moderate, three - low, four - insufficient.

We have 4 high, 11 moderate, 1 low, zero insufficient. Twenty-five percent high, 69 percent moderate, 6 percent low, zero percent insufficient.

DR. NISHIMI: Okay, overall

suitability for endorsement. Oh, I'm sorry, usability and use.

MS. CRAWFORD: Voting is open for measure 3039 usability and use. One - high, two - moderate, three - low, four - insufficient information.

We have seven high, eight moderate, one low, zero insufficient. Forty-four percent high, 50 percent moderate, 6 percent low, zero percent insufficient information.

DR. NISHIMI: Okay. Now, overall suitability for endorsement.

MS. CRAWFORD: Voting is open for measure 3039 overall suitability for endorsement. One - yes, two - no.

We have 15 yes, 1 no. Ninety-four percent yes, 6 percent no.

DR. NISHIMI: So for measure 3039 the committee recommends the measure.

The next measure I just want to say --

CO-CHAIR McINERNY: Before we leave that could we maybe ask -- just have the

committee get a sense that we would ask the CMS folks to come back in the near future with a more robust plan, or a choice of more robust plans other than just follow-up in six months.

I kind of like the idea Emilio said, a dropdown list. And I think in this day and age of EHRs there may be something that you could do with the dropdown list and hopefully you actually do what you say you did in the dropdown list. That's always a question.

And we know that I think there are -- we need to find some evidence-based measures that work, that have been shown to really work to improve the outcome.

I think we're seeing some evidence that motivational interviewing has a much better outcome than just saying come back in six months.

I don't know about referrals to dietitians, et cetera, or referrals to gyms or whatever else. I'm not sure there's much evidence for those.

But I think we need to move this up,

to raise the bar, to move this up to another level. We need to look at some evidence-based results that really help patients to improve their BMIs.

And by soon I mean a year or 2, not 5 or 10.

MEMBER TEUTSCH: I would also support Amir's contention that we really need to get this better aligned with the Preventive Service Task Force recommendation along the dimensions that he talked about, particularly about who it is that needs to get referred.

I would also point out to Tom's comment that there's been a lot of work on community-based interventions and resources.

This is not the sole providence of clinical care or public health alone. This is a real opportunity to get people connected to the resources they need regardless of whether they're inside or outside the clinical care system.

DR. NISHIMI: So we'll make sure the report reflects these.

CO-CHAIR McINERNY: Well, you can look at the community guide for one simple place, or the IOM reports on obesity control. There are a bunch of resources.

And I don't think we want to be categorical, but clinicians should be aware of what they are and take advantage of the ones that are available to them.

MEMBER STIEFEL: And especially since this also includes underweight.

MEMBER TEUTSCH: It's interesting, the Task Force doesn't talk about the underweight problem. Their recommendation is for referral of the obese.

That's not to say that you shouldn't, but it's the same issue. There isn't a specific recommendation to that effect.

DR. NISHIMI: So, I think the developer clearly heard the comments on how this measure can be improved and we'll make sure the report also reflects that.

Okay, the next measure is 2828. This

is the eMeasure version of this. So we'll carry over the evidence discussion to this measure. There's no need to revisit that.

The question is based on the information that they supplied, is there a gap. So we do need to vote on that quickly.

And then it's whether or not the testing information that they supplied, using it as an eMeasure is valid and reliable, et cetera. And then obviously the feasibility.

DR. GREEN: This is Dan Green. I'm going to drop off, but I just wanted to thank the committee. Thank you.

DR. NISHIMI: Thank you. We aren't going to discuss evidence. We're going to go straight to gap.

Is there anything new that anyone wants to add about gap? Does anyone object if we vote? Ron.

MEMBER BIALEK: And this is really a question for the developers in terms of gap.

When I was looking at the data it

seemed like you indicated that the data, it's not part of the structured field for the electronic reporting, and that also one of the three providers that you do testing with, 11 percent didn't record the BMI data.

And so I'm just -- I'm not sure, maybe I'm misinterpreting the data, but it seems to me that if it's not part of the electronic reporting how do we determine the gap.

MS. WELLS: So we tested the measure at three sites and only one of them had a dropdown menu that listed care plans such as consultation for diet, diet and exercise. And so the feasibility was very much improved obviously at that site compared to the other two.

So it speaks to the variation in feasibility I think across settings, but it is just one enhancement to the EHR. It does sort of structure that follow-up data which otherwise was in notes in care plans and sort of non-query-able fields.

In terms of the 11 percent that failed

at our one practice that did have the dropdown, I actually think what happened there was that either they didn't measure the BMI or they did and didn't provide follow-up for one reason or another.

It may have been because it was in the 25 to 30 range and they decided it wasn't necessary. So they didn't get 100 percent performance, but they did perform quite high.

CO-CHAIR McINERNY: In pediatrics almost all -- after a lot of beating of the electronic health record developers we got them to put in a system, a program that when the height and weight and gender are entered it automatically calculates the BMI and reports it. So that makes life a lot easier.

I've been around long enough to remember going around with some -- trying to do the calculations. It drove me nuts. I don't know why they came up with such an arcane formula.

But anyway, in adults I don't know if

that automatically happens.

MS. SOMPLASKY: It does not.

CO-CHAIR McINERNY: And that's certainly something that should be done is that, you know, they should know the gender. The height they can assume is staying the same. So that if you weigh them it should automatically give you the BMI. That's the first part.

Then the second part is I guess for an electronic measure the clinician, whoever it is, has to check off that they -- at this point the low bar is that they advise the patient about their being overweight and come back in six months.

But hopefully we'll get dropdown menus that will be --

MS. SOMPLASKY: And that is the variability among EHR vendors. Vendors interpreted this measure so some do have dropdown.

And one of the things we tried to do when we tested, we tested in a very small

practice, a medium-sized and a large practice with three different vendors.

Two of the three vendors were going to charge that practice to make changes. And the minute you start passing on changes like that the practices can't afford it.

So, they find -- we did find documentation in non-structured fields, but you can't pull that information to be able to provide it just from the EHR extraction.

MEMBER HARRIS: So I just wanted to -- the developer took part of the things I wanted to say is that there are certain EHRs that do have the capacity to automatically calculate.

And I think that if it's not automatically there then it's not a really good EHR system.

Secondly, I think that as a clinician and paying attention to that particular feature is going to require a little bit more than just it calculating in the background, where it will actually be prompted.

So like in the EHR system that I've recently used it actually turns the field red when you look at the actual calculations based upon what the card we used to carry in the lab coat. And we pulled it out and looked at the orange-yellow, green-yellow, whatever.

So, I'm not sure if that's in this methodology or not but that's something that is available.

CO-CHAIR McINERNY: I think as an aside we need to have the government harass EMR vendors at least as much as they harass clinicians.

MEMBER HARRIS: I concur.

DR. NISHIMI: Okay, are we ready to vote on performance gap, or any other questions? Oh, I'm sorry.

MEMBER MOLINE: First of all, I'd like to say that this measure made me feel comfortable that in a few years when I hit 65 I don't have to worry about a BMI of 25 anymore. I can go up to 30 and be considered normal so that should be

nice for everybody. Because I didn't know that. So you learn something from being on this committees.

But to your point, a lot of the EMRs do automatically calculate the BMI. And it's unusual that they don't.

At this point if the height and weight are inputted it becomes something standard. When they're inputted it becomes -- because I can tell you my staff wouldn't figure out how to do it and it's done automatically.

So at least in Allscripts it gets done.

So is there -- I think there's some other issues that we'll probably get to with usability which goes to the dropdown fields, but also goes to some of the exclusions that there was no way of getting to that we're going to get to where there's no availability in the electronic medical record to say urgent care or not, which is an exclusion. So, I think those are some more of the challenges with this

measure.

MS. SOMPLASKY: To be clear, you are correct, the BMI gets calculated. It's when it is out of normal parameters that there are a fair number of EHRs out there that do not have that information in dropdown fields so that it can be part of structured fields.

MEMBER HARRIS: Are we talking about a cookbook plan? So like if I'm 30 then you're going to do one thing, and if I'm 35 I'm going to refer you to bariatric surgery, or for 27 I'm going to just tell you to go walk to the White House and back? What exactly are we saying?

MS. SOMPLASKY: Well, that's where I said before we're not prescriptive on what that follow-up plan is.

And that is one of the things that we're trying to look at with some of those value sets.

And we've received a fair number of comments back from EHR vendors saying we need more specificity around this. So we are looking

at that for the 2018 specifications.

DR. NISHIMI: Any other comments on gap? Vote.

MS. CRAWFORD: All right. Voting is open for measure 2828 on performance gap. One - high, two - moderate, three - low, four - insufficient.

We have seven high, seven moderate, one low. Forty-seven percent high, 47 percent moderate, 7 percent low, zero percent insufficient.

DR. NISHIMI: Okay, so it passes gap and we'll move onto testing.

Again, this was an eMeasure. Any comments from the reviewers on testing?

MEMBER MOLINE: So this is where some of these are challenging. And the EMR people, the practices and the data that were submitted noted some of the challenges, that there's no way of noting whether someone's in palliative care.

More importantly if they were for an urgent or emergent visit in which case they're

not in there. And so the EMR did not allow them to distinguish that.

Some EMRs -- I mean, I'm trying to think of my various choices. It's not going to give those choices.

So as an eMeasure with these exclusions I don't know how that's going to be captured appropriately to change the numerator -- or adjust the numerator down without the EMR being modified in some way, and I don't see that happening.

DR. NISHIMI: So, you would say that the validity and reliability are affected because of this.

MEMBER MOLINE: I think they will be affected by this, yes. I mean, it's not going to be as good as it would have been with these various exclusions as described in the measure.

DR. NISHIMI: Marcel or Ron, any other comments? Anyone else, comments? Anyone else have comments? Validity and reliability of the eMeasure version.

Jacki has indicated that she has some concerns.

Okay, let's vote on reliability. It is eligible for high.

MS. CRAWFORD: We have 1 high, 13 moderate, 1 low, zero insufficient. Seven percent high, 87 percent moderate, 7 percent low, zero percent insufficient.

DR. NISHIMI: Okay, in terms of validity this is more towards where the specifications issue and the availability of the capacity to capture the exclusions.

As Jacki pointed out she has some concerns about for the empirical testing for the validity here was done at the data element level. So it will only be eligible for a moderate rating.

Any other discussion? Jacki's given her input. Marcel or Ron, did you have any additional comments? No?

Okay, then we'll go ahead and vote on validity.

MS. CRAWFORD: Voting is open on measure 2828 on validity. One - moderate, two - low, three - insufficient.

We have seven moderate, eight low. Forty-seven percent moderate, 53 percent low.

So the committee did not reach consensus on the validity criterion. We'll discuss feasibility, and usability and use, but we won't vote on overall endorsement at this time.

Feasibility. They did the NQF feasibility score card.

It noted that just as Jacki said follow-up is not in a structured field so that identifying those that fall within and outside of the measure could be problematic.

Same issue goes to the EMR version's feasibility. Any other questions?

CO-CHAIR McINERNY: Well, since I retired I don't have to deal with some of these issues anymore, particularly ICD-10.

When ICD-10 -- are there different

codes for the BMI so that if your BMI is within normal range that's one code, or if it's elevated, say a 30 it's another, and if it's 35 it's yet another.

There are. So that would make it easy to collect that measure. Then I guess the other question, and this may be more with CPT codes, if you recommend something is there a CPT code for a referral to another source or anything like that.

MS. SOMPLASKY: We actually created that on the claims registry side, to be able to have a HCPCS code to show it which is why you saw better performance there.

It's not there, and that is one of the things I was saying earlier that we are trying to look at, trying to put together value sets that would reflect that referral to another provider for a specific reason, not just in general.

To be able to get more specific for this measure. And that is something that we're working on now for 2018.

CO-CHAIR McINERNY: Thank you.

DR. NISHIMI: Okay, let's go ahead and vote on feasibility.

MS. CRAWFORD: Voting is open for measure 2828 on feasibility. One - high, two - moderate, three - low, four - insufficient.

MEMBER HARRIS: Tom, while the votes are being cast that particular ICD-10 is easy because it's really the number, whatever your BMI is. So it's a Z code and it's Z68. And then if you're 30 it's Z68.30. If you're 27 it's dot 27.

So it's not easy to remember all the ICD-10s period, but like we couldn't remember the ICD-9s, but this is one you could remember.

DR. NISHIMI: We're still looking for one more vote. Point your clickers again and hit them again, please. We're still waiting for one more.

MS. CRAWFORD: Okay, we have six high, six moderate, three low, zero insufficient. Forty percent high, 40 percent moderate, 20 percent low, zero percent insufficient.

DR. NISHIMI: Feasibility isn't a

must-pass criterion so we'll now vote on usability and use.

Any comments from the committee on usability and use? Let's go ahead and vote then.

MS. CRAWFORD: Voting is open for measure 2828 on usability and use. One - high, two - moderate, three - low, four - insufficient information.

Two more votes please. We have four high, nine moderate, two for low, zero insufficient information.

Twenty-seven percent high, 60 percent moderate, 13 percent low, zero percent insufficient information.

DR. NISHIMI: Okay, there was a concern by a committee member that they mishit on the validity vote which was consensus not reached. So we're going to go back and redo the validity vote. They're afraid they didn't hit the correct button.

So this is validity for the eMeasure. The high comes off.

MS. CRAWFORD: So one is moderate, two is low, three is insufficient. Voting is open again for validity of measure 2828. One - moderate, two - low, three - insufficient.

Okay, we have 10 moderate, 5 low, 67 percent moderate, 33 percent low, zero percent insufficient.

DR. NISHIMI: A lot of people must have thought they -- okay, so then all the must-passes have been cleared, so now let's vote on final.

MS. CRAWFORD: Voting is open for overall suitability for endorsement on measure 2828. One - yes, two - no.

Okay, we have 14 yes, 1 no. Ninety-three percent yes, 7 percent no.

DR. NISHIMI: Okay. We have one more measure to do, 3062, Hypertension Screening for Children who are Overweight or Obese. Are the developers here? Are the developers on the phone?

So while we're waiting for the

developers we'll ask them to introduce their measure briefly.

This is a new measure. It's a plain old measure. It's not an eMeasure. So we'll just discuss as we do evidence, performance gap, testing, et cetera.

MEMBER MOLINE: This shouldn't skew. I just wanted to say that this was the first time in any measure that I've seen socioeconomic status in a measure in any of the ones that I've looked at.

And last time we were all together she was sitting over there, I forget who it was, but every single measure she said she wanted to see socioeconomic status.

So I feel like she's here in spirit. I forget what her name was. It was Renee. So, someone heard her.

MS. MUNTHALI: Is Q-METRIC or University of Michigan with us on the phone?

Okay, so let me just see who from there. Operator?

OPERATOR: Yes, ma'am.

MS. MUNTHALI: Is Gary Freed, Joyce Lee, Julie McCormick, or Caroline Shevrin on the phone from Q-METRIC or the University of Michigan?

OPERATOR: I don't see either one of those on the line yet.

MS. MUNTHALI: Interesting. Okay, so we're going to try and email them really quickly and see.

So while we do that we're going to try and get you out early. And we will forego our discussion on harmonization of the influenza measures.

We want to make sure that we have as much of the committee here, and we know you're very tired because you've done a lot of work and so we thank you so much.

So we're hoping we can get this measure reviewed as well. If not, this will have to be in the post in-person meeting call which I think is on the 22nd of this month.

So, we're going to give the developers one last chance to join us. We're not sure what happened.

CO-CHAIR McINERNY: While we're waiting for that could we have one or two of the subcommittee members give their input, like Matt or Jacqueline?

Well, yes, I'm also -- yes. Go ahead, Matt.

MEMBER STIEFEL: Sure. So, there's going to be I think a pretty significant issue with regard to evidence so this could be a short discussion in that the evidence provided isn't specific to the performance measure because the evidence isn't specific to this testing above the 85th percentile for BMI for hypertension.

So the staff review concluded it was insufficient and I concur with the staff review.

DR. NISHIMI: Jacki, you were also a reviewer. Was that your conclusion on the evidence?

Matt's conclusion was that the

evidence was insufficient.

CO-CHAIR McINERNY: Yes. And I also agree.

DR. NISHIMI: So, the question I guess to Matt and Tom and Jacki is would you recommend it for the exception so that it would be a national performance measure, or should the committee just vote.

MEMBER STIEFEL: I think I would not because there's -- I think if you go through the list of the subsequent categories there are also problems with a number of the other elements including the second one in terms of the opportunity for improvement.

The opportunity for improvement with one study showed that there was opportunity, but the reliability and other validity measures were -- the information presented was not sufficient in my review.

MEMBER MOLINE: I felt it was an immature measure in that it didn't have enough data behind it to be able to show evidence.

There's not -- there's some good evidence and there's not. And ironically the New York Times talks about blood pressure screening in kids today so that high-quality journal is talking about the importance. So I thought that was very topical for today.

But apart from that they just didn't -- the measure developers didn't present enough information for us to be able to adequately evaluate.

And I think they need a couple of more years or more data to be able to come back and present it.

I don't think it's a bad thing to present, but I just think there wasn't enough there yet.

DR. NISHIMI: Yes, they had a small sample size that once they kept extracting they got down to six records.

MEMBER STIEFEL: You're right, the review came down to six records.

DR. NISHIMI: Okay. I'm going to do

this. We're still trying to get a hold of the developer. We'll allow them to make their pitch if they come on before we're ready to leave.

But let's vote on evidence.

MS. CRAWFORD: Voting is open on measure 3062 on evidence. One - high, two - moderate, three - low, four - insufficient.

Okay, we have zero high, zero moderate, 1 low, 13 insufficient.

DR. NISHIMI: So then let's go to the insufficient with exception vote just to close it out.

So this is whether the committee wants to consider giving this measure the exception.

MS. CRAWFORD: The options are one - insufficient evidence with exception, two - no exception.

We have 1 for insufficient evidence with exception, 13 with no exception.

DR. NISHIMI: Okay, so the measure is not recommended by the committee.

Elisa, do you want to take it? Tom?

CO-CHAIR McINERNY: Let me just -- eventually when you report back to the developers my notion is that this is a good idea, we really should be doing this, it's just not quite ready yet for prime time I guess.

And they need to work better on the measures and so forth.

And one of the things, I don't know if, again because I haven't been in practice in two or three years, whether EMRs are doing a better job of calculating whether the blood pressure is greater than the 95th percentile.

For children it's not an easy calculation. You need the gender, you need the height and you need the blood pressure. So that's a three-parameter table.

And it covers about -- for males it covers two pages and for females it covers another two pages.

So, to do it by hand you get the blood pressure reading and then you've got to go through this.

Now, I've talked to some people who say that the EMR should be able to calculate that just the way they can calculate a BMI. The parameters are there, the height's there, the gender's there, the blood pressure is there so the EMR should be able to say whether it's above the 95th percentile or not.

But I'm not sure that happens in many, if any EMRs and I think that's part of the thing we should ask the developers to let us know about whether EMRs are beginning to do this in a more regular fashion.

It's a lot more difficult than adults because of the height is an important factor.

Plus the age. I should say age is another factor. So it's height, age, gender, blood pressure. Those four parameters have got to be all taken into account.

DR. NISHIMI: Thanks, Tom.

CO-CHAIR McINERNY: Thank you.

DR. NISHIMI: Elisa?

MS. MUNTHALI: So, we've completed a

review of 24 measures over 2 days. That's quite a bit. A number of eMeasures, composites, intermediate outcome measure, an outcome measure, a number of influenza measures.

And we were due to talk about harmonization of those influenza measures. And we'll do that later.

But we really wanted to thank you so much for getting through all of the agenda items.

You've probably worked enough. We will not I think -- we're not going to have the post-meeting call. So we're going to give you back time. We'll cancel it on your calendars.

But before I turn it over to public comment I just wanted to thank each and every one of you. We know that the turnaround time was very short, the volume of material you received was a lot and we just wanted to thank you so much for your thoughtful review.

We thought it was such a successful meeting. We want to thank the developers and also our co-chairs Tom and Amir today for

facilitating such a wonderful meeting.

And I'd like to thank my colleagues as well. So with that I'll turn it over to our operator to see if there are any public comments on the phone and then see if anybody who's left, if there are any public comments here.

So, Operator, can you open up the lines?

OPERATOR: Yes, ma'am. At this time if you would like to make a public comment please press star then the number 1. And there are no public comments at this time.

MS. MUNTHALI: Thank you very much. And I'll turn it over to Yetunde for next steps.

MS. OGUNGBEMI: Good afternoon. I thank you for a very fruitful and full discussion today and yesterday.

I want to review the next steps of the committee and basically the remaining timeline.

On well, I should say later this week staff will begin drafting the final report. And this will be complete with committee

deliberations.

And we will post this draft report for NQF member and public comment from October 20 through November 18.

Following that commenting period the committee will meet to review and discuss measures where consensus not reached. And that is on December 6.

NQF member vote will happen from December 21 to January 4, 2017 on recommendations by the committee.

The Consensus Standards Approval Committee, warmly known as CSAC, will meet in either January or February of 2017 to review committee deliberations and recommendations made on a date to be determined. And that should be coming in soon weeks.

The committee co-chairs will be asked to join that call and provide CSAC with specific nuance if needed.

And recommendations finally will go through appeals from February 3 to March 16.

DR. NISHIMI: I just wanted to lend my thanks to everyone. This was a heroic effort. Twenty-four measures is really at the outer limit, and you really marched through them I thought with very thoughtful discussions.

As the staff drafts the report we might be in touch with you. You know, we said this. We aren't quite sure what exactly you meant, or we want to make sure we capture the appropriate nuance.

But other than that there won't really be any major follow-up till you get the comments and decide on those consensus not reached measures.

CO-CHAIR McINERNY: And on my part and I think for the committee's part we couldn't have done the 24 measures without your help.

Your evaluations helped guide us and made I think the process run more smoothly.

Now, on this calendar I'm a little confused. What's on September 22? Anything?

DR. NISHIMI: That's the thing that

we'll send you the cancellation for.

CO-CHAIR McINERNY: Okay. And then on December 6, that's still on the calendar?

DR. NISHIMI: Yes, you should all have that in your Outlook appointments.

CO-CHAIR McINERNY: Okay.

DR. NISHIMI: So what happens is we'll get the comments in and we'll let you know what people think.

CO-CHAIR McINERNY: Good. Okay. Thank you very much.

MS. MUNTHALI: Thank you. Travel safely.

DR. NISHIMI: Safe travels to all.

(Whereupon, the above-entitled matter went off the record at 2:59 p.m.)

H

M

Q

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Health and Well-Being
Standing Committee 2015-2017

Before: NQF

Date: 09-13-16

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Court Reporter