NATIONAL QUALITY FORUM

+ + + + +

INFECTIOUS DISEASE ENDORSEMENT MAINTENANCE STEERING COMMITTEE

+ + + + +

TUESDAY
AUGUST 28, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Steven Brotman and Edward Septimus, Co-chairs, presiding.

PRESENT:

STEVEN BROTMAN, MD, JD, Advanced Medical Technology, Co-Chair
EDWARD SEPTIMUS, MD, FACP, FIDSA, FSHEA, HCA Healthcare System, Co-Chair
JEFFREY BEAL, MD, AAHIVS (via telephone)
MARY BLANK, MPH, CIC, CPHQ, Highmark, Inc.
KATHLEEN BRADY, MD, Philadelphia Department of Public Health
DOUG CAMPOS-OUTCALT, MD, MPA, University of Arizona, Phoenix
RAYMOND CHUNG, MD, Massachusetts General Hospital
CURTIS COLLINS, PharmD, MS, BCPS, University of Michigan Health System
SUE ELAM, BSN, PHN, MHS, FNP, Kaiser Permanente Medical Group
MOHAMAD FAKIH, MD, MPH, St. John Hospital and Medical Center
MICHAEL C. FARBER, MD, Department of Vermont Health Access

THOMAS M. FILE, JR., MD, Msc, MACP, FIDSA
THOMAS GIORDANO, MD, MPH, Harris County Hospital District
PETER HAVENS, MD, MS
AARON MILSTONE, MD, MHS, Johns Hopkins Hospital
REKHA MURTHY, MD, FRCP8, FACP, Cedars Sinai Medical Center
TIFFANY OSBORN, MD, MPH, FACEP, Washington University/Barnes-Jewish Hospital
KALPANA RAMIAH, DrPH, MPH, Msc, CHES, CPH, CTTS, American Institutes for Research
DAVID SPACH, MD, Harborview Medical Center
ADAM THOMPSON, Consulting

NQF STAFF:

HEIDI BOSSLEY
HELEN BURSTIN
ANN HAMMERSMITH
ADEELA KAHN
NICHOLE McELVEEN
ALEXIS MORGAN
REVA WINKLER

ALSO PRESENT:

JEFFREY CLYMAN, Resolution Health, Inc.
BEN HAMLIN, National Committee for Quality Assurance (via telephone)
EMANUEL RIVERS, Henry Ford Health System
JOHN WONG, Tufts Medical Center

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

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

(8:59 a.m.)

DR. WINKLER: Good morning, everyone. I am Reva Winkler. I am the Senior Director of Performance Measures here at NQF. Thank you all for being with us today.

As we get started, I would like to introduce our project team. I think we are all names that you are familiar with over email, but now we have faces to put to them. So over sitting at the table near the window is Project Manager Alexis Morgan, and with her is our Project Analyst Adeela Kahn.

Sitting next to me is the Senior Vice President for Performance Measures, Dr. Helen Burstin. Did you want to say anything?

DR. BURSTIN: No.

DR. WINKLER: The Co-Chairs for this committee, sitting next to me, are Dr. Ed Septimus and Dr. Steven Brotman. So we need to get to know everybody else on the committee well. So to lead the introductions and

disclosures, I would like to introduce NQF's General Counsel, Ann Hammersmith over here in the corner, and I will let Ann tell you what we need for us to do for introductions.

MS. HAMMERSMITH: Good morning, everyone. As Reva said, we are going to combine introductions with disclosures in the intro. So what we will do is we will go around the table. You introduce yourself, tell us who you are with, and let us know if you have any disclosures that you would like to make.

To refresh your memory of that disclosure, several months ago you received a fairly detailed form from us where we asked you a lot of information about you and your professional activities. We reviewed those, and the analysis of the disclosures is a component of what we use to select members for the committee.

So what we would like you to do is to disclose anything that you think is relevant to what is before the committee during this

meeting. It doesn't mean you have to disclose your full CV. Please don't; we will be here all day. We know you are all extremely competent, and that is why you are on the committee.

We do ask you to disclose just what is relevant to what is before the committee. Just because you disclose doesn't mean you have a conflict. It just means that you are letting your fellow members know about you and your activities.

We are particularly interested in your disclosure of grants, research or consulting activities that may be relevant to what is before the committee. I also want to remind all of you that you serve as an individual. You are not a representative of your employer. You are not representing the interests of anyone who may have nominated you for service on the committee.

Sometimes we have committee members very innocently say I am Suzie Jones, and I am

here representing the interests of the American Association of -- fill in the blank. Actually, you are not. You are here as an individual expert.

The last thing I want to remind you of is that your interests can be other than strictly financial. Members will sometimes say I have no financial conflict of interest. Because of the nature of the work in this field, there may be something relevant that you need to disclose where you weren't paid for it. You may have been a volunteer on a committee where the work of the committee might be relevant to what is before the committee.

So with that, I am going to start with the Chairs, and we can go around the room.

CHAIR BROTMAN: I am Steve Brotman from the Advanced Medical Technology Association, known also as AdvaMed. I don't have any disclosures to make.

CHAIR SEPTIMUS: Ed Septimus. Good morning. I am the Medical Director of

Infection Prevention and Epidemiology at HCA in Nashville and have an academic appointment at Texas A&M Health Science Center in Houston, and I have no relevant disclosures for our discussions.

MEMBER THOMPSON: Good morning. My name is Adam Thompson. I am a person living with HIV, a patient at Ryan White Care, and I am a consultant with the National Quality Center, a grantee of the Health Resources and Services Administration.

MEMBER RAMIAH: Hello. I am Kalpana Ramiah. I am a principle project specialist with American Students for Research and also adjunct faculty at George Washington University. No disclosures to make. Thank you.

MEMBER FARBER: Hello. I am Michael Farber. I am a full time employee of the University of Vermont, College of Medicine, and I serve as the Vermont Medicaid Medical Director. I have no relevant disclosures.

MEMBER FAKIH: I am Mohamad Fakih. I am Medical Director of Infection Prevention at St. John Hospital Medical Center. I also serve as a physician for infection at Ascension Health. I am supported partially by HRAT, which is the arm of the American Hospital Association for the national work on catheter-associated urinary tract infection.

MEMBER SPACH: I am David Spach, based at Harborview Medical Center and have an academic appointment to the University of Washington, and I have no relevant disclosures.

MEMBER GIORDANO: Good morning. I am Tom Giordano. I am at Baylor College of Medicine in Houston. I also have an appointment at the Houston VA Medical Center, and I have -- I am Medical Director for an HIV clinic called Thomas Street Clinic that has substantial Ryan White funding. I have also done contract, consulting and grant work with HRSA, NIH and CDC.

MEMBER CAMPOS-OUTCALT: Doug

Campos-Outcalt with the University of Arizona College of Medicine Phoenix campus, and I currently serve on two panels, one the Advisory Committee on Immunization Practices, and a second is EGAPP Working Group of CDC.

MEMBER HAVENS: I am Peter Havens at the Medical College of Wisconsin and Children's Hospital Wisconsin in Milwaukee, Wisconsin. I have done contract work in the area of HIV with CDC, HRSA, and I get research funding from NIH.

MEMBER COLLINS: Hi, good morning. I am Curtis Collins. I am a clinical pharmacist with the University of Michigan Health System. Conflicts: I am a member of the American Society of Health System Pharmacists Council on Therapeutics, as well as the society of Infectious Disease Pharmacists Public Policy Committee. No relevant financial conflicts.

CHAIR SEPTIMUS: The Chair forgot to say that he is also an Ohio State Buckeye,

but we will not hold that against you.

MEMBER COLLINS: And that is true.

MEMBER BLANK: Good morning. My name is Mary Blank. I am from Highmark, Blue Cross/Blue Shield, in Pittsburgh, Pennsylvania, and we are an insurance company, and I manage and oversee the development of programs that are designed to improve health care quality, a number of pay for performance programs. I have no financial conflict of interest.

We do use many of NQF endorsed measures in our program models. Thank you.

MEMBER ELAM: Good morning. I am Sue Elam. I am a family nurse practitioner, and I work at Kaiser Permanente in Sacramento, and I work in the Department of Infectious Diseases, the HIV Care Clinic.

MEMBER BRADY: Hi. I am Kathleen Brady. I am the Medical Director, Medical Epidemiologist for the AIDS Office for the Philadelphia Department of Public Health where I receive multiple grants through CDC and work

on quality management projects for our Ryan White programs for Part A and Part B.

I am an infectious disease physician at Pennsylvania Hospital, which is part of the University of Pennsylvania Health System, and in terms of financial disclosures I am on the speakers bureau for Gilead Sciences.

MEMBER MILSTONE: Good morning. My name is Aaron Milstone. I am on the faculty of Pediatrics at Johns Hopkins University. I am an infectious disease consultant at Johns Hopkins Hospital, and I am also one of the Associate Hospital Epidemiologists.

In terms of disclosures, I am a co-director of infection control at Kennedy Krieger Institute, across the street from Johns Hopkins. I have NIH grant support to look at strategies to reduce catheter-associated blood infectious. I have received a research grant from Sage Products, similar to look at an intervention to reduce catheter-associated infections, and also I have some leadership

positions at the Society for Healthcare Epidemiology of America.

MEMBER FILE: Good morning. I am Tom File. I am infectious disease clinician in Akron, Ohio. I am Chair of the Division of Infectious Disease at Summa Health System in Akron and Chair of the Infectious Disease Section at Northeast Ohio Medical University.

I think the only relevant disclosure may be that I authored the section and up-to-date on acute bronchitis, which we will be discussing, but in light of the comments from our Co-Chair, I will disclose also for his benefit that I did graduate from the University of Michigan Medical School.

MEMBER MURTHY: Good morning. I am Rekha Murthy. I am at Cedars Sinai Medical Center as hospital epidemiologist and at the faculty in the Infectious Diseases Division, and have a faculty appointment at UCLA at David Geffen School of Medicine. I have no relevant disclosures for today.

MEMBER CHUNG: Hi. I am Ray Chung. I am Director of Hepatology, the lone hepatology wolf in this room, I suspect, and have grant funding from the NIH, and an officer with the American Association for the Study of Liver Diseases and have conducted clinical trials for a number of companies, including Gilead, Roche, Merck and Romark.

MEMBER OSBORN: Well, I am another lone wolf, I think. So my name is Tiffany Osborn, and I am an attending physician at Washington University. Half my clinical time is in the emergency department, and half my clinical time is in the surgical trauma intensive care unit.

My disclosures are relevant to the topic that I will be presenting, which is I have been a representative from the American College of Emergency Physicians to the Surviving Sepsis campaign for over a decade. Additionally, I have worked with the Institute of Healthcare Improvement to assist them as a sepsis

consultant in sepsis measures where they are doing locally determined variation of early goal directed therapy within a hospital system, and I am the trial clinician for ProMISe, which is protocolized management in sepsis, which is evaluating early goal directed therapy for severe sepsis in sepsis shock within the United Kingdom involving around 48 sites.

MS. HAMMERSMITH: Okay. I understand there is one committee member on the phone.

CHAIR SEPTIMUS: Jeff?

MEMBER BEAL: Yes, thank you. Hi. I am Jeffrey Beal. I am with the Florida Department of Health. I am the Medical Director of the HIV/AIDS and Hepatitis Program, and I am also the principal investigator and Clinical Director of the Florida Caribbean AIDS Education Training Center, and I have no financial disclosures. Thank you.

CHAIR SEPTIMUS: Jeff, we are sorry you can't be here, but I understand there was

a storm in Florida.

MEMBER BEAL: Yes. Actually, we got off easy. It is very wet and windy, but unfortunately, my flight was canceled, and as a DOH employee we embargoed from travel at times of potential disasters. So I am sorry I cannot be there in person, and thank you for understanding.

CHAIR SEPTIMUS: We are glad that you are safe.

Just a few comments, if I can, before I turn it over to someone else. Oh, I am sorry. Excuse me.

MS HAMMERSMITH: Just one little wind-up piece. Thank you for the disclosures. Do you have any questions of me or anything that you want to discuss with each other based upon the disclosures this morning? Okay, thank you.

CHAIR SEPTIMUS: Okay. Let the people know that we are ahead of schedule, and we hope we continue that way.

I am going to turn this over to Dr. Winkler in just a moment to really go through some of the details, which some of you may have seen, but I think will be helpful for this morning's discussion. Before that, just a couple of things.

We want to keep everything on time. We want to be respectful, and we want to make sure that the measures at the end of the day get the same focus that the measures do at the beginning of the day.

The way these meetings have worked best: If you have a comment you would like to make, if you will just turn your name tag sideways, we will keep track of who wants to comment. I think that is a nicer way to do that. When I think everything is said that needs to be said, we will try to move it along to voting, but we want everything that needs to be said to be said.

After we comment, each of you have taken a measure that you will present to the

group, and there will be comments from the group. As you know, we have a time for public comment as well, which will occur after the discussion.

You all have received these little clickers here. So keep them handy for voting, and we will go through this in detail, but when we get to the votes and the different levels, we will be using this to vote, and then our votes will be tabulated, and then we will move on with the discussion.

With that, Dr. Winkler has got some really important slides she wants to go over with us, some of which you have already seen, but I think will set the stage about the order in which these measures will be voted upon. Reva.

DR. WINKLER: Thank you, Ed, very much. I wanted to review sort of the context of the work you are doing and how it fits into the big picture of particularly what NQF does and the meaning of NQF endorsed measures.

NQF, I think you are all well aware

of, is a private nonprofit organization, but it is a public/private partnership and very much a multi-stakeholder organization. Member organizations represent the wide spectrum of stakeholders, including consumers and purchasers, as well as professionals, providers, community public health, measurement folks, research folks, health plans, supplier and industry.

So the members of this committee are a proxy for that very diverse membership, and so we do have deliberately people on this committee who bring different perspectives. One of the great values of NQF is to be able to share those different perspectives. With that, we hope all of you will feel comfortable offering your thoughts and sharing your perspective with your colleagues.

NQF has several missions. Building consensus on priorities and goals is something that happens primarily in our Strategic Partnership Division, but what we are most known

for and have been doing for the 11 years of NQF's existence is endorsing national consensus standards for measuring and publicly reporting on performance.

That is essentially the work you are doing. You are helping us do the evaluation of candidate measures to be endorsed by NQF for use in public reporting and other accountability purposes. This is sort of NQF's foundational work. So we do thank you very much for being part of it.

NQF's role is as a standard setting organization. As such, we do endorse voluntary consensus standards in the areas of performance measures, the serious reportable events, some preferred practices and frameworks. Today we are looking at performance measures. But NQF also, particularly in the last five years or so, has expanded its work as a neutral convener of several other important collaborative efforts.

One of them is the National

Priorities Partnership, which is a collaborative of 51 major national organizations which brings together public and private sector stakeholders to balance all of those interests. Probably one of their noteworthy activities is provide direct input to the Secretary of HHS on the National Quality Strategy.

So the National Priorities Partnership is an ongoing enterprise within NQF. Another NQF convened partnership collaborative is Measures Application Partnership, again another multi-stakeholder group, that provides input to HHS on measures that should be used within the Federal programs.

So both of those groups rely very heavily on the performance measures that NQF endorses in the Performance Measures Division.

Now the National Quality Strategy was announced a little over a year ago by the Secretary of Health and Human Services, and NQF's work is geared to support the National

Quality Strategy of better care, healthy people and affordable care.

So those principles in the NQS really do reflect patient centeredness, quality of care, elimination of disparities, and alignment of public and private sectors. So it is important to understand how the work we are doing will contribute to the National Quality Strategy, and specifically we have measures in this project that do look at promoting better care, most effective treatments for infectious disease specifically around HIV, hepatitis C and sepsis.

Also, we are looking at affordable care or appropriate care, you might say, specifically around the overuse of antibiotics. We will be talking about disparities for each measure, and determining whether there is a characteristic disparitive sensitivity to each of these measures, and then we are looking to align public and private sector work for all patients with these conditions.

So why NQF endorsement? What is the point? A couple of them, actually: Standardized performance measures are the tools that can be used to assess quality on a national basis, that can be used to make comparisons, and they need to be good enough for that purpose. So please keep that in mind.

NQF endorsement reflects a rigorous assessment, evidence based review, input from all of the different stakeholders and perspectives throughout the health care industry.

So as we look at the measure evaluation criteria, please be aware that that criteria has evolved over time to reflect the input of a wide variety of stakeholders and the needs that those stakeholders have voiced in terms of measures that are going to be used to hold people accountable for the care that they deliver.

NQF endorsed measures are widely used. We have over 700 measures in the

portfolio, and this was an analysis we did at the very beginning of 2012 looking at how NQF measures are used. You can see that about half of them are used in Federal programs. Additionally, others are used in states or by private payers. Then there are some other additional uses. Only a very small percentage, around six percent, we were not able to determine that they were currently in use in a major program.

So you can see that that is why we are here and how these measures are used. The current infectious disease measures are used in many of these programs, specifically Medicare's Physician Quality Reporting System, which is a physician or clinician level accountability program; NCQA HEDIS measures for some of these measures; and then several states are using some of the measures in their enterprises. Then many others are using them for quality improvement. So for the most part, these measures are used by a large number of

organizations.

The endorsement maintenance process, which you are a critical part of, is to ensure the currency and relevance of NQF's portfolio of measures, and for this care in the area of infectious disease.

It is our goal to review measures that have been endorsed by NQF every three years. However, we do do an annual update to determine if there have been any changes to the measures or any changes to the literature or evidence or anything that would promote a more earlier review.

So the majority of the measures that are before you, all but five, are measures that have been previously endorsed by NQF. Be aware, however, that over the time NQF's processes have evolved. Our measure evaluation criteria have become more specific and perhaps set a higher bar for measures. So simply because a measure was previously endorsed does not necessarily mean it would meet the current criteria.

In terms of the endorsement maintenance process, we solicit measures, new measures, to be brought into the process, as well as identify those measures in this topic area that are due for a maintenance review.

We also seek implementation comments from the field asking how is it going out there with these measures; are there particular problems with implementation; is there something you can offer to share with us in terms of how it is going.

All of the measures, whether maintenance or new, are reviewed against the same criteria with the same expectations of meeting those criteria.

We also, after we have reviewed all the measures, will be looking at measures that seem to be similar or addressing similar topics, looking to see if the measures really are harmonized in the way the definitions, in the way they look at measures, to make it easier for those in the field to be able to implement

the measures; and if there are measures that are essentially identical or so similar as not to matter, then perhaps we need to discuss whether one can be chosen over the others.

So this is our process. Schematically, the Steering Committee is the pivotal committee to review. You are acting as a proxy for NQF's membership, multi-stakeholder, varied perspectives. You are going to do the first initial review of the measures against the criteria.

After that, your recommendations will be put out for public comment. We will get comments from NQF members and the public on your recommendations, and then this group will regroup to look at those comments to see if they may -- that feedback changes any of your thoughts about the measures.

Once you get a chance to review those and rethink based on the comments, those become draft consensus standards. They go to the NQF membership for voting.

The voting results go to our Consensus Standards Approval Committee, the CSAC, which is a subcommittee of the Board whose specific task is to oversee this consensus development process, and then finally ratification by the Board of Directors. That ratification grants the NQF endorsement, and then there is a 30-day appeals period.

So this is a very formal process that is meant to achieve consensus in a structured fashion.

The other thing I wanted to mention is we will be asking about disparities. That is some of the information we request on the submission forms. We would also ask any of you with your expertise and experience if you can offer additional information so that we can better understand disparities-sensitive measures and identify those within our portfolio, for those folks who particularly want to focus their measurement activities around reduction of disparities.

We do have a protocol around disparity sensitive measures. I would like to introduce my colleague, Nicole McElveen, who is sitting next to Alexis, who leads our disparities work. She and her Steering Committee have identified a method to look at disparity-sensitive measures that are focused around the prevalence of the condition of minority populations, the disparities quality gap which is completely dependent on data.

This is where some of our biggest struggle is. It is not necessarily having the data we need to really understand how significant or how big a quality gap may exist in disadvantaged populations; also whether it rates high on impact, particularly pertaining to the National Quality Strategy, and then whether it maps to an NQF preferred practice from the communications or care coordination domain. That doesn't seem to apply to many of the measures in this particular project, though it certainly does in others.

So we will be talking about disparities-sensitive measures.

So that is sort of the overview, and I would like to give anybody an opportunity to ask any questions to be sure you all understand why you are doing what you are doing today, the big picture of how this contributes to the overall quality measurement enterprise and how the results of your work might be used going forward.

CHAIR SEPTIMUS: Tom?

MEMBER GIORDANO: Just a quick question, Reva. When you say disparities, you distinctly mean racial/ethnic disparities, not disparities based on income, gender, sexual orientation, anything else?

DR. WINKLER: All of those would be appropriate. So it is not restricted to race and ethnicity. Helen, did you want to add?

DR. BURSTIN: Although at this point, unfortunately, most of the data we have available is based on racial and ethnic

minorities and disparities therein. If there are additional data to be brought to bear, I think the same process and algorithm that our committee has come up with would still work.

DR. WINKLER: So I would say you wouldn't want to restrict it, but I think our information is limited on which to know much about it. Anything from anyone else? Anybody on the phone? Dr. Beal, did you have any questions?

MEMBER BEAL: No, ma'am. Thank you.

DR. WINKLER: Thanks.

MEMBER FILE: Thanks for that very nice overview, Reva, but just one quick question. What input does NQF have on pay for performance initiatives?

DR. BURSTIN: The way this is currently organized is that the endorsement process is really looking at the measures themselves. Do they meet our criteria? Are they measures that are reliable, valid,

important, etcetera, evidence based, that could be used?

We then have a separate entity called the Measures Application Partnership that NQF is the organizer for, but it is really a group of external folks all coming together, multi-stakeholder. Part of their role is every winter they get a list of all the measures proposed by CMS for all the programs, and they make specific recommendations. That is separate and apart from this.

So there may be some of the measures that you are looking at that may wind up being in payment. Some may wind up being used primarily for QI in the interim. Some may be used for benchmarking and other purposes.

I think at this point what we would ask you to do is stay somewhat agnostic of how they will be used, and instead focus on the quality of those measures themselves. Again, you need to, at the same time, though, I think, consider that any of the measures you are putting

forward could be used for any of those potential applications, should another group agree that that is appropriate.

DR. WINKLER: The other thing I would add is the MAP only looks at the public sector, but there are lots of pay for performance programs in the private sector who use a wide variety of measures. So NQF endorsement is a source of measures that a wide variety of organizations look to, to put into their various programs.

CHAIR SEPTIMUS: Aaron?

MEMBER MILSTONE: Just to follow up on that, there were some comments in the Work Group summaries. It seems like it is clear that some measures may be good for internal QI, but once you start to look at them across institutions or across states, there's going to be lots of differences.

So I wonder, how should we factor that in when we think of good for internal QI versus bad for general comparisons of practices

across the country?

DR. BURSTIN: It is a great question. In general, measures that are really only appropriate for internal QI don't rise to the level of being endorsed by NQF. So there may be -- There are thousands of measures, as we all know, that people are using out there for the sake of internal QI. We want to have measures that rise to the level of you are comfortable that you actually can do valid comparisons, have important information available for consumers and those who purchase on their behalf to make valid decisions.

DR. WINKLER: Now to the work at hand for today, and that is the evaluation of the measures before us. You all have had an opportunity to look at the measures. You have all had the opportunity to participate in the Work Group conversation. So this is really the end of what has been a process over the last several weeks of evaluating these measures against the criteria.

What we have provided for you to help is a summary of all of that work. We have given you a hard copy. We sent you the electronic one on Friday. This is the summary of all of your preliminary submissions. It is also our best summary of your discussion. So this is sort of the jump-off spot for your conversation today.

Also at your places we have given you a four-pager, I think it is, three- or four-pager that is quick guide to the evaluation criteria. You have seen that evaluation criteria in so many different ways and shapes and forms. We are hoping at least one of them will resonate with each of you.

I do want to hit some highlights to begin with, and for our first measure, please bear with us. What I would like to do is to, as Dr. File goes through the evaluation, just review the criteria with you, with the first measure, to allow you to be sure we are focusing in on the right thing.

Just to give you a background of these evaluation criteria, we really have created criteria to help -- to ask the questions to be sure that the measures do meet the criteria that the stakeholders have determined are important for using these measures.

So the subcriteria under each of the main four criteria demonstrate how those criteria are met. So the questions are around how do you know a measure is important; how do you know that a measure is scientifically acceptable.

We believe that these criteria have been developed because they parallel the best practices for measure development. Measure development should start with good evidence base and a good development of measure specifications and then testing of those for reliability and validity.

Most of the criteria, however, are just not black and white. So we wouldn't need you, if they were. So we really are looking

to your expertise and experience, either in the clinical area or in measurement or in some other aspect of the quality enterprise, to provide that extra subjective overlay that is needed to really assess these measures and the information provided against our criteria.

So new versus endorsed measures: As I mentioned, we really -- Everybody is expected to meet the same measure. So one of the criteria for ongoing endorsement is not that it was previously endorsed. That is just very straightforward.

So we really are, though, however, looking for information on those endorsed measures on how it is going out there, data from current use, current implementation. I think there is a question to be raised that, if there is no data, why not? Is it not being used and, if so, why not? But also reliability testing: We are hoping that, again, ongoing use, more and more reliability and validity assessments are done so we can really understand how solid

these measures are.

Usability: I think this speaks to the question that Aaron may have just asked, actually use in public reporting or other accountability activity, or specific plans for it. We are not looking to endorse measures that are intended and will only be used for internal quality improvement. That is really, actually, a pretty hard break point.

Then feasibility: Can it be done? What do we know about it in terms of data sources, data collection, data crunching? What do we know about it? So these are really important information, particularly on previously endorsed measures that we may not yet have for new measures.

Just another reminder: We have shared this information with you in the staff memo, but the CSAC looks at all of the measures in the portfolio over and over again, and they come up with some sort of themes of things that just don't really seem to work very well for

many of the stakeholders, and you all have brought up some of these issues, but it is worth repeating because they will push back if measures are brought to them.

They are not particularly encouraging measures that can be simply met through documentation, the checkbox measures, if you will. Also, the fact that teaching and counseling should be viewed from the patient's perspective to determine how effective that teaching and counseling was.

Consider the impact of missing data. Excluding missing data can really be problematic in calculating reliable and valid measures.

The exclusion should be evidence based or sufficient frequency that it will really impact the results. Measures should be specified with the broadest applicability, populations such as applying to children when appropriate. Can we use the same measure in the inpatient/outpatient post-acute care

setting? Levels of analysis, when appropriate.

Then look at how the measure is constructed, and avoid measures that, as you get better, the denominator gets smaller. We have seen some of those, because they become difficult measures to handle as improvement occurs.

So these are just some basic guidelines that the CSAC wants you to be aware of in terms of the kinds of measures that they are looking to put in NQF's portfolio.

So now as we get closer to actually getting to work, we have asked each of you to take on the role as the lead discussant for each measure. You have had an opportunity to do that in the Work Group. It is a way of sharing the work around the table and getting everybody to contribute.

I would ask each of you as you introduce your measure to declare the name, title, and read the description of the measure so everybody kind of is on the same page and

knows what we are talking about. Also, it helps people who may be listening on the phone to know what we are talking a bout.

Then we are going to go through each of the subcriteria and criteria that need to be voted on one at a time. So I would ask the lead discussant to summarize your thoughts and that of the Work Group discussion on how well the measure and the information provided to you about the measure meets or does not meet that NQF criterion. That is the fundamental question before you as we go through this multiple times today.

After the lead discussant gives you that intro and summary, everyone on the committee is encouraged to offer your thoughts, ask questions, clarify, because we need you to be able to comfortable rate the measure on that criteria.

Since you have probably only look in detail at the measures in your Work Group, you are relying on each other to share the

information from the other Work Groups so that you can serve your role as a full committee member for all of the measures. So the entire committee will vote on to what degree the measures meet all of the NQF criteria.

So are there any questions about the role of the lead discussant?

CHAIR SEPTIMUS: Go ahead.

MEMBER FAKIH: I have a question regarding process versus outcome measures. if you have a well established outcome measure, would you accept also a process measure? Let's say you have a very validated outcome measure at present? What is the role of the process measures in that case?

DR. BURSTIN: That is a great question and one I don't think there is a clear answer to. I think, in general, we have a hierarchical preference in the way we have looked at measures for outcome of a process. If they are going to be process measures, they shouldn't be very distal, far away from the

outcome. They should the ones most proximal to the outcome. Assessment measures, for example, probably don't have a place if you have proximal measures closer to the outcome that are really more meaningful.

At the same time, it is oftentimes very useful, particularly for those who are being measured, to have a suite of measures that allow them to see what is potentially impacting on the outcome. So we wouldn't necessarily exclude the process measures as long as they are, in fact, quite proximal to the outcome. So that is a decision you are going to have to talk through.

CHAIR SEPTIMUS: And correct me if I am wrong. Most of the measures we are going to look at are going to be process measures, because one of the challenges with outcome measures is that they have to be risk adjusted, and that gets to be, for many of these measures, difficult.

DR. WINKLER: Just to remind you,

we will be going through the four major endorsement criteria. Under Importance, we will be asking you to on the three subcriteria of importance, evidence, and opportunity.

Under Scientific Acceptability, we will be asking you to vote on reliability and validity, and then a vote on usability, feasibility, and then an overall vote on suitability for endorsement.

Now I also think -- Each of you have been given a little voting keypad, and with the first measure we will have a chance to check it out, but be sure. Does everybody have one? Okay, because using this we will be able to see the votes up on the screen and see how things go.

I just want to remind you that the importance to measure and report is a must-pass criteria. So if the measure fails on any of the subcriteria, we stop at that point. So that is why it is important that we capture the votes in a real time fashion.

Similarly with scientific acceptability for either reliability or validity -- If they don't pass, that's it. We stop. Usability and feasibility are not required to be passed, and so the committee will use their judgment in determining if they have an issue with usability or feasibility, whether it is overall suitability for endorsement.

So with that, one thing under importance is I am going to ask your indulgence. We have discovered over the course of several projects that the discussion seems to go better if we reorder it and talk about impact first, evidence second, and opportunity third.

So we haven't reordered the numbers. So it is going to seem a little strange that we will go from A to C to B, but I think we can all cope with that. So I just wanted to let you know that that is where we are at.

Also, at the beginning of each either group of measures or measure, we are going to give the measure developer an opportunity

to introduce their group of measures. Most of them are fairly grouped, so that there won't be an introduction prior to each and every measure but around their group of measures in that particular topic.

I think we are probably ready to get started.

CHAIR SEPTIMUS: Okay, have your quick guide out. I think I have found this to be terrifically useful for the voting purposes and where the stop points are, similar to what Reva just went over, but it is nice to have the quick guide out.

Secondly, I think I am also going to assume that all of you are studious folks and have had some -- gained some familiarity. I have to commend, by the way. The NQF staff is absolutely incredible, and I want to thank each of them individually for the incredible amount of work they did and how fast they came up with the summary of our call. So I thank Reva and her staff very much for this. We

couldn't do this without your support. So we thank you very much.

We are going to assume that you have some familiarity, even if you were not in the work Group where the calls took place. So, Tom, you are going to lead the first one. So assume that some of us have at least familiarized ourselves with the measure. So we don't have to go through every single detail that was provided.

So we have the developer for this measure. Would they like to speak first?

DR. WINKLER: Ben, are you on the phone?

MR. HAMLIN: Yes, I am. Good morning. My name is Ben Hamlin. I am the Director of Performance Measurement for NCQA. My comments are regarding 0058, Avoidance of Antibiotic Treatment in Adults With Acute Bronchitis, and 0069, Appropriate Treatment for Children With Upper Respiratory Infection.

These are two measures, both of

which are being currently used in HEDIS, are both measures in the PQRS measure list. Both measures were included in the MPRM. However, only the URI measure made it to the final rule due to concerns about burden.

They are both effectively measures that address overuse of antibiotics, one obviously in adult and one in children, and they are very, very similar in their approach. They are reported at an inverted rate. So the higher rate indicates better performance, so, therefore, indicating the appropriate use of antibiotics in these two populations.

I will leave my comments at that.

CHAIR BROTMAN: I think, Aaron Milstone, you have a question, please?

MEMBER MILSTONE: We were just trying to -- Just before we start, could you give us some information just about the voting? Is this a majority vote?

DR. WINKLER: Yes, it will be a majority vote for each of the subcriteria.

CHAIR BROTMAN: Tom, did you want to start?

MEMBER FILE: Yes, I would be happy to. So the first one is 0058, Avoidance of Antibiotic Treatment in Adults With Acute Bronchitis. It is a maintenance review endorsement. The description: Assesses the percentage of adults ages 18 through 64 years of age with a diagnosis of acute bronchitis who are not dispensed an antibiotic prescription.

Some additional comments by the developer are that the IDSA Quality Improvement Task Force endorses this, as well as 0069.

I guess we will just start out with the first, which is the importance -- or the impact, I'm sorry. I think it is fairly well consensus anyway that there is overuse of antibiotics in this particular diagnosis. The diagnosis is a very common one presenting to ambulatory centers and emergency departments.

It is also known from a variety of studies that at least 90 percent of these

infections are due to vital etiology, for which the use of antimicrobial, or at least antibiotics, would not be warranted, and would not be beneficial for the patient.

As a matter of fact, as we know, the use of antibiotics in these types of conditions are a significant harm in that it increases the selection of resistance for the common pathogens, and we have all too well seen what that has done in the last couple of decades.

So from the standpoint of impact, I am not sure -- I will be happy to entertain any comments. Do you want a vote?

CHAIR SEPTIMUS: Do you want impact first? We have on the lefthand side the measure report. Before we vote, are there any questions from the group before we vote on the impact? Okay, then we will just move forward.

As I understand it, your last vote is the one that counts. Is that right? You can change your mind?

DR. WINKLER: If you change your

mind, you know, the last vote is what counts. Adeela's computer has the receiver. As the countdown starts, we will be able to see how many people have voted. So when we reach the point where everybody has voted, we will be able to stop it and show the results. So why don't we use this as sort of our first pass.

So, Adeela, are you ready to go?

CHAIR SEPTIMUS: One question. You push the number and Send? Just the number, as I understand. That is what I understood. I want to make sure that it is not confusing.

DR. WINKLER: Just the number. Ready to go?

MS. KAHN: We are going to vote on 1(a), High Impact: Addresses a specific national health goal or priority, and the data demonstrated a high impact effect of health care.

CHAIR SEPTIMUS: Is your mic on?

MS. KAHN: Sorry. So you want to press 1 for High, 2 for Moderate, 3 for Low,

and 4 for Insufficient, and you can go ahead and start voting.

CHAIR SEPTIMUS: That was a really close vote. Okay, Jeff? High, Moderate, Low, or Insufficient?

MS. KAHN: So we have 19 votes for High, zero for Moderate, zero for Low, and zero for Insufficient.

CHAIR SEPTIMUS: Just to let the committee know that, since this is a public meeting and there are people on the phone, we have to verbalize and repeat all of it. So just to let you know why we do that.

Tom, do you want to talk about the evidence?

MEMBER FILE: Okay, for the evidence -- Now let me just clarify. Do you want me to go through each of the three subcategories of evidence first; so then we go to quantity first?

DR. WINKLER: I think, if you can summarize them together, that is fine. That

way, you hit each of those points.

MEMBER FILE: So we would then vote on evidence as a total.

DR. WINKLER: Yes.

MEMBER FILE: Fine. As far as the evidence, we are fortunate in this respect, that there are several systematic reviews, and most recently just earlier this year, there was a Cochrane Systematic Review that was published. It was actually performed last year, 2011, which is an update of a prior Cochrane Review.

The most recent review evaluated 15 trials, which is increased from the prior Cochrane Review, which comprised 2,618 patients, and I am just going to quote from that review that they found that there was limited evidence for any marginal effect of antimicrobials. However, the magnitude of a small benefit needs to be considered in the broader context of potential side effects, increased resistance, and cost of the antimicrobial treatment.

Their conclusion was this update provides clear evidence on the lack of effectiveness of antibiotics for acute bronchitis. So the fact that we have got the systematic review, I would just refer to that for the evidence or at least a quantity of evidence, 15 -- These are randomized clinical trials. Fourteen of them were placebo, double blind, randomized clinical trials. So theoretically, level 1 evidence.

As far as the quality, again if you read that Cochrane Systematic Review, they evaluate for consistency -- Well, that is the third. They evaluate for consistency then, and selective bias, and heterogeneity, and found that these pass those criteria. So at least from the standpoint of the Cochrane Systematic Review, they felt that the quality was adequate.

As far as consistency, if you look again at that review, there was a very -- or a consistent pattern of results from these studies. Most of them, about 12 or 13, did show,

depending on what their outcome was, a very minimal potential benefit. For example, number of days of cough, for example, may have been reduced by .5.

I think the issue, when you look at all these studies, is the enrollment criteria, in that most of them did not require a chest X-ray to rule out pneumonia. So there may have been some patients enrolled in these studies that would have benefitted from some antibiotics, because they may have had pneumonia.

Nonetheless, then when you looked at the studies that evaluated potential adverse events, obviously, the placebo won there. So that is sort of a basic review of those 15 trials at least that were included in that systematic review.

CHAIR SEPTIMUS: Any questions or comments regarding the evidence? If that is not the case, then could we put up the voting slide?

DR. WINKLER: I want to point out in the voting slide that you have three voting options. One is Yes, it meets the criteria for quality, quantity and consistency. There are two types of No votes. One is that the evidence does not meet those criteria, and 3 is there is insufficient information to know whether they meet the criteria, given the information presented to you. So you do have those two options for No.

MS. KAHN: Voting on 18, evidence. We are looking for a rationale that, based on information submitted, the quantity, quality and consistency of the body of evidence are met as follows: The consistency is Moderate or High, and the quantity and quality are Moderate and High, or Low with special circumstances.

So you are going to vote 1 for Yes, body evidence meets the guidance for quantity, quality, and consistency; 2, No, evidence does not meet the guidance for quality, quantity and consistency, including no empirical evidence

exists; and 3, No, Insufficient Information submitted to raise the quantity, quality and consistency of the body of evidence. So you can begin your vote.

We have 19 votes for Yes, the body of evidence meets the guidance for quantity, quality, and consistency.

CHAIR SEPTIMUS: Jeff?

MS. KAHN: Oh, we got his vote. So there are votes for No, evidence does not meet the guidance; and there are votes for No, there is insufficient information submitted.

CHAIR SEPTIMUS: Now the next one is opportunity.

MEMBER FILE: The opportunity, I think, is very clear when you look at the performance gap, but at least as illustrated by that that was presented by the developer from information from the HEDIS data collection, which indicated over the last three years anywhere from like a 25 to 22 percent that actually met this measure. So that the

majority, almost 75 percent, did not meet the measure.

So, obviously, there is a significant room for improvement, based on that information. In fact, then if you look at a variety of observational studies, it shows similar information -- or results, I should say.

CHAIR SEPTIMUS: Any comments or questions on opportunity? Jeff, since you are on the phone, too?

DR. WINKLER: One question.

CHAIR SEPTIMUS: Go ahead, please.

MEMBER HAVENS: It is Peter Havens. I interpreted this as 23 percent -- Since this is one minus the rate. So I thought 75 percent met; 23 percent get antibiotics inappropriately. Do I misunderstand the measure?

MR. HAMLIN: Yes. The initial assessment was correct, that the 23 percent does indicate the appropriate performance.

CHAIR SEPTIMUS: Correct.

DR. WINKLER: Do we know anything about disparities for this kind of issue?

MEMBER FILE: Well, the developer -- I will refer to NCQA representative, but at least in the application they say that there is no strategy for that except using ZIP Codes.

DR. WINKLER: Okay. Do you all have any sense of whether, for this particular process of care, disparities are an issue?

MR. HAMLIN: This is Ben from NCQA. We do continue to look at the availability of disparities information for our different HEDIS measures. Unfortunately, the data is not consistent enough for us to make any kind of assessment at this point in time.

We do see a variation in rates across the different regions, and we have a variety of different theories as to why that is, but we don't have a specific ZIP Code analysis for the disparities at this time.

CHAIR SEPTIMUS: Tom, a question?

MEMBER GIORDANO: Could you just

clarify whether -- to follow up on Peter's question -- is it 75 percent have met the standard and are not dispensing antibiotics?

MEMBER FILE: No. Twenty-two percent meet. It is just the opposite.

MEMBER GIORDANO: Thank you.

MEMBER FILE: Seventy-five percent of patients get antibiotics for 466.0, at least according to their data.

MEMBER FAKIH: Just a comment. You know, first this is a coding that we are going to be tracking. So it is coded data, which may be -- There may be a shift in diagnosis through coding data with acute bronchitis. So that is one of the worries that I would have also for this measure.

The other point is that what Dr. File has raised, for the last three years there was no improvement, although it has been adopted by certain groups as a quality measure, but there was no improvement.

So although there is a huge gap, but

this may not -- Having it as a measure, I am not sure it will affect this rate to change.

CHAIR SEPTIMUS: Let me comment. There is always a concern that people are going to code to justify the use of an intervention, and I think that is certainly something we would certainly look at.

Secondly, I think the reason that this has not budged very much is there is very little accountability for not doing the right thing, and I think until we have some accountability with organizations, we have a very slow improvement. I think it has to do with accountability.

MR. HAMLIN: Yes. And to address your coding question, there's two things. So for the HEDIS data, auditors must sign off on the results that are submitted by the health plans, and they do look for shifts in measure rates, and they would go back and look and see if there was a major shift in coding practices.

We are also investigating different

ways right now that we can look and see the frequency of the codes used to identify certain conditions. We are going to try and identify several databases where we can try and get a better understanding of that. That was also driven by results of the Work Group feedback that we received.

CHAIR SEPTIMUS: Any other questions? I don't see any. So are we ready to vote on this measure? Okay.

MS. KAHN: Voting on 1b, performance gap: The data demonstrated considerable variation and overall less than optimal performance across providers and/or population groups, and we are looking at disparities in care. Vote 1 for High, two for Moderate, three for Low, and four for Insufficient Information. You can start voting.

So we have -- do some quick math -- 16 votes for High; two votes for Moderate; zero for Low; and one Insufficient Evidence.

DR. WINKLER: Adeela, is there some way we can make the projection show the vote count rather than the percentages?

MS. KAHN: Yes. I am going to change it right now, actually, before we go on to scientific acceptability.

CHAIR SEPTIMUS: You have to put on your microphone.

MEMBER RAMIAH: Sorry for that. Total number is always 19, but I thought we had 19 here, with one person on the phone.

MS. KAHN: Dr. Beal is putting his votes into our webinar. So we are using a clicker to capture his vote.

MEMBER RAMIAH: So it is 19.

CHAIR SEPTIMUS: Total of 19.

MEMBER RAMIAH: Yes.

CHAIR SEPTIMUS: Okay. It is time for Reliability.

MEMBER FILE: Again, just for clarification, do we vote both of these together or separate?

DR. WINKLER: You vote for first reliability, then validity.

MEMBER FILE: Well, then first is reliability. According to the application and at least for the criteria for reliability, in fact, that it is well defined as is specified, it is well defined and specified in that you are looking at a specific ICD-9 Code for 66.0.

The developers provided information of a reliability calculation using HEDIS health plan performance data, reported both from, I think, Medicare or Medicaid and then a commercial database of .96 and -- Well, actually, it was .96 and .99 respectively. So it does suggest that this is a valid -- excuse me, a reliable, at least repeatable, measure.

Now I will comment. Mohamad said that, if one concerns the potential -- and Ed addressed this as well -- of changing in codes, that is one thing, but if one looks at specifically what this measure is to do, and that is measure 466, then it is very

straightforward.

DR. WINKLER: I just want to remind us on the criteria for evaluating and rating reliability and validity. We are looking for empiric testing. Testing can be done at the data element level or at the measure score level, and in this case it appears to be done at the measure score level with a type of signal-to-noise analysis.

If it has only been tested at one of the two levels, the highest rating you can give it is a Moderate. All right? So at this point, a High rating on reliability doesn't mean it is highly reliable. It means -- you are talking about NQF's criteria, and our criteria for High means you have to have tested it at the data element level and at the level of the measure score. So I just wanted to remind you of that.

Moderate is it will be fine enough to pass, but realize that that is what the criteria is. Alexis, can you go one more. Keep

going. There you go. So High is, note, only if tested at both levels. Moderate can be tested at either level, and the results are good, obviously, as well as the precision specification. So I just wanted to remind everybody that this is the rating scale for reliability, similarly for validity, to make sure we are all kind of on the same page for the evaluation.

MEMBER FILE: Well, let me just ask so I am sort of clear on this: Obviously, the developers provided a measure score. Now as far as data elements, the data element would be assessing for the specific ICD-9 code and how they actually are able to determine that.

DR. WINKLER: Perhaps. typically, the kind of testing of the data elements are around the specific either codes in the numerator or the denominator, the critical data elements, and the kind of empiric testing we typically see is inter-rater reliability, particularly if they are abstracted and whether

they are abstracted in a similar fashion.

So that is looking at the individual elements of the measure. Testing at the level of the measure score is what you are seeing here, the results and whether the signal-to-noise analysis.

So you can see, and you can look at the reliability of the measure at both levels. So the criteria for a High rating is that it has been measured at both levels, and it comes up High. Okay? This is a clarification for everybody about the criteria.

MEMBER FILE: I guess I am still not clear. Is there in the application evidence of measure at the data elements? I mean, I should be telling us this, but I want to make sure.

DR. WINKLER: I don't believe there is.

MEMBER FILE: I agree.

MR. HAMLIN: We do not. We didn't include the original field testing data that

was accomplished in 2003, but we are happy to provide that information, if you feel it would help your decisions.

CHAIR SEPTIMUS: Peter?

MEMBER HAVENS: So the initial review committee seemed divided on this issue fairly evenly, if I understand the format here. Could we get some input to the larger group on how they sorted that out in their discussion, since it is a three and three split on reliability, and by your estimation the data are not included. So that would suggest that they don't pass, unless I --

CHAIR SEPTIMUS: No, this --

MEMBER HAVENS: That could be Moderate if the data are not included?

CHAIR SEPTIMUS: Peter, if you look at it, it was split between High and Moderate.

MEMBER HAVENS: Yes.

MEMBER FILE: And I can tell you -- and I hope I am representing our group accurately -- that much of this is based on, really, more

of a concern for feasibility or unintended consequence of changing in codes, and I am going to report that when we talk about feasibility. But if you look at our comments there, a lot of it had to do with the fact that there was a concern for the fact that there was a shift in coding.

CHAIR SEPTIMUS: Any other comments on reliability? Okay, then we will vote.

MS. KAHN: Voting on 2a, Reliability. Includes 2a1, precise specifications, and 2a2, testing the appropriate method and scope with adequate results. So you are going to vote 1 for High, 2 for Moderate, 3 for Low, and 4 for Insufficient. You can begin your vote now.

So we have two votes for High; 15 for Moderate; one for Low; and one Insufficient Evidence.

DR. WINKLER: The majority are High or Moderate, and that is sufficient.

CHAIR BROTMAN: Okay. The next

one, I believe, is validity.

MEMBER FILE: Next is validity, and correct me if I am wrong, Reva. here we are going to really look at does this truly in a valid way measure the discrimination of the performance; that is, those who have a poor or those who have a good performance of this measure.

The developer presents a fairly extensive process with a variety of committees and experts in this and a public reporting and review to support this measure as being valid as to being able to differentiate poor from good performance of this measure.

DR. WINKLER: I think you are describing face validity as opposed to empiric testing of validity.

MEMBER FILE: Yes. Thank you. I am not sure what I was presenting, but I appreciate that interpretation.

DR. WINKLER: The important thing about face validity, again because it is not

empiric testing and is only face validity at the highest level, you should rate that as Moderate. But also we need to talk about any potential threats to validity, and I think this is where your coding issue might come up.

MEMBER FILE: Right, and that is where the one that ranked low, I think, was the point, was the concern about the coding issue.

DR. WINKLER: Would you like to share that a little bit more with everybody? I am not sure everybody got --

MEMBER FILE: Well, I can talk about it now, but to me it really is more of a concern for an unintended consequence when we see these measures put into practice. In fact, there was a study -- two of us actually brought this out during the discussion -- that was just published two months ago in the American Journal of Managed Care or Clinical Journal of Managed Care, whatever that is, but at any rate, it looked at a health care plan database from the years

2006 to 2009 as far as the response to 466, which is the code for acute bronchitis, and found that in this particular health plan there was a significant reduction in the use of antibiotics for this code.

On the other hand, they also observed a significant shift from 266 to 490, which is bronchitis not otherwise specified. When you looked at the combined effect of 266 and 490, there was just a minimal or a marginal, perhaps modest at the most, reduction in antibiotic use, and they suggested that the influence of a measure to reduce antibiotics in 466 led to many prescribers using a different code to justify the use of antimicrobial agents.

So that is something that we discussed during our workshop with the developer and, as already has been discussed, they are looking at shifts in particular health care plans to see if that is a trend, but that really was the concern of the validity here.

CHAIR SEPTIMUS: Does the developer

want to make any comments, and then we will ask for questions?

MR. HAMLIN: Sure. So again, our initial field testing was looking at 466 and 490 to look at the prescribing rates by diagnosis code, and the initial testing across four plans' different claims' diagnosis indicated using multiple claims to ID both the diagnosis and comorbidities between the two, that the use of 466 was the appropriate code and the use of 490 was the inappropriate code. However, again that information is from 2004 and, therefore, this is why we are going to go back, in light of the new evidence, and investigate how to retest this to ensure that those findings are, in fact, consistent in a larger database across the nation, across different plans.

CHAIR SEPTIMUS: Question?

MEMBER FAKIH: This may be a question for the developer. So this is again a coding diagnosis. It may not be reflective of what the physician has written in the chart,

but it is what was billed for. Is there a way that we in the future, if we have this as a measure, to figure out if this is really what the physician or the provider has put as a diagnosis? How can we reconcile the coding to the true diagnosis that the physician has entered, or is this something we worry about, because this, I think, will hurt validity quite a bit.

MR. HAMLIN: That issue is getting a lot of scrutiny right now for the meaningful use Stage 2 measures, and there was an extraordinary amount of attention paid to the different diagnosis codes across value such that were used to identify the denominator and the numerator for these measures.

I think that that will probably be something that we will be looking at when we get to validity and reliability testing of those measures. Right now, we have only accomplished feasibility testing for the EHR measures.

CHAIR SEPTIMUS: Any other

questions regarding validity? Please.

MEMBER THOMPSON: Did that paper you referenced -- did it look at -- So another possibility, rather than codes are shifting in appropriately, is that codes are shifting appropriately, that some of those people -- you know, it was easy to say acute bronchitis when there were no consequences, and you just kind of tagged it as that, before this measure was adopted. Now people are coding more appropriately when it is not acute bronchitis.

Did the paper try to distinguish those two possibilities?

MEMBER FILE: Well, I will have to look at this in closer detail. My recollection is no. It was just sort of an observation and a suggestion that there was an influence in the measure that altered the pattern of the coding, but I will look at that in closer detail.

CHAIR SEPTIMUS: Any other questions?

MEMBER THOMPSON: I just want to

make sure that I am reading this correctly then. Based on the submission that I see here -- and I am just looking at the notes -- were threats to validity assessed in the submission form?

MEMBER FILE: Well, the threats by what criteria are you looking at?

MEMBER THOMPSON: I am just looking. So one of the things it is saying, that the threats were empirically assessed in biased results, and under potential threats to validity, there is "Not Applicable" on it. So I am just making sure that I grade it appropriately, like were there threats to validity and, if so, were they assessed in the submission form?

MEMBER FILE: I think the threats that we have discussed were those that we were concerned about, primarily related to coding.

MEMBER THOMPSON: Okay.

CHAIR BROTMAN: Kathleen.

MEMBER BRADY: You are asking my question, but they are not addressed in the

submission form.

MEMBER FILE: Correct. That is correct.

DR. BURSTIN: I think what Ben was telling us is they have field data. It is somewhat outdated from 2003 in which this was assessed. I think what he is saying now is, given the shift to EHRs, they are now going to be looking at those threats more significantly in, I think, a more appropriate platform of EHRs.

MR. HAMLIN: Yes, that is correct. We did do a thorough analysis of 466 and 490, and at that point in time in 2004 there was no substantial impact on the overall measure rate, but given now that we are now into ICD-10 and SNOMED coding diagnosis in the EHR measures, we are going to be doing a thorough analysis to determine which codes are being mapped within the current EHR systems for these measures, since they are both in the NPRM.

DR. BURSTIN: We should probably get that information, Ben -- this is Helen --

sent to us, just so we have it for completeness.

MR. HAMLIN: I am happy to provide that report.

MEMBER BRADY: In terms of our voting, we are voting based on the discussion rather than what is in the submission form, or vice versa?

DR. WINKLER: I think that, particularly since Ben is going to add some of that information into the form, you can factor in the discussion. That is the purpose of it.

MEMBER BRADY: Okay.

CHAIR BROTMAN: Peter?

MEMBER HAVENS: No, that was my question as well. We have been given specific instructions to assess the data that are on the forms that we were given, and it should be fairly straightforward to see where these are measured and, if it is not, then it is difficult to know exactly how to vote in a reproducible manner, if the information is supposed to be on the form.

MR. HAMLIN: The reason we did not

include the information at this point in time, as I said, the field testing was done 10 years ago. We rely on our audit process to identify any major shifts in the measure rates, which would then indicate that there is a shift in the use of these diagnosis codes. As well, we also have -- we rely on our software certification vendors to examine any kind of major shifts in coding that might affect the rate. But we are certainly interested in retesting this information as it was brought to us that it may be an issue here.

MEMBER HAVENS: Then the vote would seem to me to have to be Not Available. I need some feedback here just to understand. I am asking for guidance from you guys, because I am new to the process, and it is hard to know when everybody says vote based on what is supplied in the paper, and then we hear about stuff that isn't supplied.

DR. BURSTIN: I think that is a very fair question. I think, if you look at what

is listed there, you know, with the exception, I think, of very significant details on threats to validity, there is a description on the actual submission form of what they did around testing for threats to validity. I think at that point --

MEMBER HAVENS: But there is no actual results. The description says what they looked at, but when you say -- I am just looking at the 0058 outline, trying to get a summary of what is currently available. I would be glad to have it pointed out where I can say this is High, so that I could understand exactly how to make this decision. From the primary review committee would be fine. You guys looked at this in some detail. I have reviewed this and your comments. The group seemed to be pretty evenly divided. So I am trying to understand how to interpret it. That's all.

CHAIR BROTMAN: Tom, do you want to address this?

MEMBER FILE: Actually, if you look at validity, I mean the majority had it high, but actually in retrospect, based on these criteria that may be adjusted somewhat. But as I interpreted it, if you really look at the measure, it is looking at a specific ICD-9 code.

If you just look at that measure, what they are measuring of 466, to me, it is very valid, because it does differentiate poor from good performance. The issue is, to me, more one of feasibility, when we get to that of unintended consequence, that we have observed with the changing of the ICD-9 code patterns.

So if you just specifically look at the measure, which is just looking at 466, to me, it is not that much of an issue as far as validity. I think I would appreciate other interpretations from our developers, but I guess that is how I was sort of interpreting it.

CHAIR SEPTIMUS: I think -- Oh, Aaron.

MEMBER MILSTONE: Just to follow

up, I guess to understand validity, I also interpret validity as how well does that ICD-9 coded of 466.0 identify patients with acute bronchitis. That is not just feasibility. That is also a validity thing, is how valid is that in correctly identifying the population of patients of interest.

MEMBER FILE: Yes. Well, that is true.

MEMBER HAVENS: So where are the data presented that show that it does that? That is my question. There are no data that I see presented here that say it does that. The question was raised already on the other side of the table.

This makes good sense to me as a great measure, but if we are supposed to be using criterion based votes, I don't see where these criteria are laid out and these questions that Mohamad raised earlier are answered in the data.

CHAIR BROTMAN: Mohamad, go ahead.

MEMBER FAKIH: Twenty seconds. I

think this is a major threat for the validity of this measure. So if we code correctly what we are seeing, then there is no issue, but if we don't code it correctly, there is a huge threat for validity. From my standpoint, I am going to vote depending on how I feel about that coding, whether it is accurate or not. I don't think there is any additional information.

There are some -- You know, there is the publication Dr. File talked about that shows that you can have a shifting diagnosis. Now how often this happens, I don't know.

MR. HAMLIN: Yes. I would like to offer, if I may, that we haven't got done this detailed analysis, because apart from several observations there may be a shifting in diagnosis, there really isn't a lot of evidence to indicate that there is.

So, therefore, when the issue was brought to us that this may be an increased concern now, that we are going to now investigate it, but again the initial testing information

was that there really wasn't.

DR. BURSTIN: Ben, this is Helen. Can you just provide for us a verbal assessment of what the 2003 field testing showed, at least to give some sense of what the results were to the committee? Do you have that in front of you?

MR. HAMLIN: Sure.

DR. BURSTIN: Thank you.

MR. HAMLIN: Sort of the denominator population, the percentage of denominator that was entered by the use of 466 was between 77 and 81 percent across different plans. Percentage of 499 was 18 to, it looks like, 25 percent, so an average of about 22 percent.

So there was roughly an average of a 22 percent reduction in the denominator for the use of 490 and, given that, our expert panel at that time suggested that it exclude 490, because they were concerned about the unspecified designation of that diagnosis.

However, they did feel that, by only including 466, they were capturing the proportion of population that did have, in fact, acute bronchitis.

CHAIR BROTMAN: At this point, I think we need to move on, and we are going to move to the vote on validity.

MS. KAHN: Voting on 2b, validity, including 2b1, specifications are consistent with the evidence; 2b2, the testing is appropriate method and scope with adequate results and threats; 2b3, exclusions; 2b4, risk adjustment and stratification; 2b5, meaningful differences; 2b6 comparability and data sources.

So you are going to vote 1 for High, 2 for Moderate, 3 for Low, and 4 Insufficient Information, and you can start voting now.

We are going to try that one more time, actually. We have an extra vote. So you can begin now.

We have zero for High; 11 Moderate;

1 Low; and 7 Insufficient Evidence.

CHAIR SEPTIMUS: Well, this measure then does pass. This is one of the stop measures, by the way. So this one does pass. So now we are going to go one to usability. So we want to keep moving, because we have one more measure before we take our break, but I think you are getting the hang of this.

MEMBER FILE: For usability, criteria is meaningful, understandable, and useful to the intended audience, public reporting, and quality improvement.

I think, from the evidence that we have seen, there is, obviously, room for improvement in this particular measure. it is used for public reporting by certain health care plans. Obviously, this is highly recommended by a variety of public policy organizations.

Again, as you can see, there were comments from our group about this issue of appropriate coding and how valid that is that we have just discussed.

CHAIR SEPTIMUS: Any additional questions on this? I think this one is a little more straightforward. If there are no other questions, we will go on to vote.

MS. KAHN: Voting on usability: 3a, meaningful, understandable, and useful for public reporting and accountability; and 3b, meaningful, understandable, and useful for quality improvement. You are going to vote 1 for High; 2 for Moderate; 3 for Low; and 4, Insufficient Information. You can start voting now.

We have 9 High, 10 Moderate, zero Low, and zero Insufficient Information.

CHAIR SEPTIMUS: Now we are going to go to feasibility.

MEMBER FILE: Feasibility then: The criteria is clinical data generated during care process or electronic data. Susceptibility to inaccuracies or unintended consequences, then data collection strategy can be implemented.

Now right now this is based primarily -- and our developer can correct me if this is inaccurate -- based on billings, but they are going to be transitioning to EHR. Again, this is where the issue, I think, of unintended consequence may play its biggest role, and that would be 4c, at least based on that one paper that we discussed.

CHAIR BROTMAN: Yes?

MEMBER GIORDANO: A lot of exclusions from the denominator. You have to search for antibiotics in this case. It seems very cumbersome, just looking at it. Can the developer comment or anyone here comment on whether this is something that is useful in the field?

MR. HAMLIN: Yes. It is an administrative claims measure only, and we do include a number of codes to identify comorbid conditions where the use of antibiotics might, in fact, be appropriate.

The reason is to sort of create

almost an exception rule to make sure that the provider is not being unfairly dinged for the appropriate use of antibiotics. But since it is administrative claims, the programming is done through certified software vendors' administrative claims algorithm that looks for these different comorbid conditions within a certain time frame from the initial encounter and diagnosis.

CHAIR BROTMAN: Yes, Mary?

MEMBER BLANK: I would like to comment that we use this measure in our pay for performance programs for physicians and patients at our medical home models, and it works very well from a claims assessment type of methodology.

CHAIR BROTMAN; thank you. Any other questions?

CHAIR SEPTIMUS: I will just mention from this standpoint, those practices that are on EMR, we are already capturing this information and feeding it back to the

physician. So in terms of feasibility, it is feasible.

CHAIR BROTMAN: All right. If there is no more discussion, let's go to voting on feasibility.

MS. KAHN: Voting on feasibility: 4a, the data are generated during care; 4b, electronic sources; 4c, susceptibility to inaccuracies and unintended consequences are identified; and 4d, data collection can be implemented. So you are going to vote 1 for High, 2 for Moderate, 3 for Low and 4 Insufficient Information. You can start voting now.

I think we are missing one person. So if you could all press your response again.

We have 8 High, 10 Moderate, 1 Low, and zero Insufficient.

CHAIR SEPTIMUS: So the last one is the overall suitability for endorsement. Obviously, this is another one of those stop measures. If you don't endorse it, it doesn't

go. Is there any other discussion? I think we are ready to vote on this. Seeing no comments, let's go ahead and vote on the suitability for endorsement.

MS. KAHN: Voting on overall suitability for endorsement: Does the measure meet NQF criteria for endorsement? Vote 1 for Yes and 2 for No, and you can start voting now.

We have 19 Yes, and zero No. So the measure will pass.

CHAIR SEPTIMUS: Thank you, Tom, and the developer. We are going to try to pick up some speed and go on to the next measure, which has a lot of overlap with this measure. We may be a few minutes late for break, but I still want to make sure this measure gets the same consideration.

Who is going to do this? Okay, Rekha is going to do this measure. Thank you.

MEMBER MURTHY: Thank you. I think, as you already mentioned, there is a lot of overlap with this. This one is number 0069,

appropriate treatment for children with upper respiratory infection.

The data reflects percentage of children three months to 18 years with a diagnosis of URI who are not dispensed antibiotic, similar to the adults. So many of the issues are very similar, but if we move right to impact perhaps, again the issues of antibiotic resistance as well as adverse events as a direct correlation to unnecessary and overuse of antibiotic utilization are applicable in this population as well and, in particular,a because of the number of upper respiratory illnesses on average for children under the age of five is more frequent than with adults.

So, certainly, the importance of the topic has been addressed through multiple studies, as shown in the citations. In addition, there is the Cochrane Review that also reviewed and concluded on the importance of addressing antibiotic overuse.

I think those are the main points.

CHAIR SEPTIMUS: Okay. Does anybody have any questions? I think this is one we can probably vote on fairly quickly, unless there is a question. Well, let's vote.

MS. KAHN: Voting on 1a, high impact. Vote 1 for High, 2 for Moderate, 3 for Low, and 4 Insufficient Evidence. You can start now.

I think we are missing one person, if you could all enter your response one more time. We have 19 High, one Moderate, zero Low and zero Insufficient Evidence.

CHAIR SEPTIMUS: We shouldn't have 20. I thought we were at 19. This is not Chicago, folks.

MEMBER MURTHY: Twenty is correct now.

CHAIR SEPTIMUS: Okay. All right, the next one is going to be evidence.

MEMBER MURTHY: Again addressing the evidence supporting the measure includes

the six trials with a total of 1,047 participants, randomized trials comparing antibiotic therapy against placebo, demonstrating again, I think, complicitly the importance of unnecessary use of antibiotics in this particular setting.

So the evidence hasn't been graded, but it was thought to be high enough for a guideline to be developed. I think that is probably all we need for this Do we have any comments on it? There is a lot of overlap.

CHAIR BROTMAN: Any discussion?

DR. WINKLER: Just one question in terms of the criteria. Do you have enough information to assess the quality, quantity, and consistency of the evidence?

MEMBER MURTHY: I think it is the -- Again, there are more studies in adults than in children, but I think there is a lot of corollary findings, and I think in terms of the developer's assessment, there was moderate quantity, quality, and consistency of the

evidence. I think that would be an accurate assessment, I think, from our standpoint and also from the standpoint of the Work Group call.

CHAIR BROTMAN: If there is no more discussion, let's vote on the evidence.

MS. KAHN: Voting on 1c, evidence. Vote 1 for Yes, the body of evidence meets the guidance for quantity, quality, and consistency; 2, No, evidence does not meet the guidance for quality, quantity, and consistency; and 3, No, Insufficient Information submitted to rate the quantity, quality, and consistency of the body of evidence. You can start voting now. Again, we are looking for 20 votes. If we are not there, just keep clicking.

We have 15 for Yes, the body of evidence meets the guidance; 3 for No, the evidence does not meet the guidance; and 2 for No, there is insufficient information submitted.

CHAIR BROTMAN: Let's move on to

opportunity and performance gap.

MEMBER MURTHY: In terms of the opportunity and performance gap, certainly, pediatricians do much better than adults in terms of avoiding antibiotic use. Looks like in this particular one. There is data from two different sources from 2009 to '11, show roughly 83 to 85 percent are not dispensed and allowed. So about 15 percent meet the measure, and that is a big difference from adults. However, the opportunity, I think, still remains with the millions of doses of antibiotics that are probably unnecessary.

On the other hand, it does look like there has really not been a big movement in this, just as with the adults, in spite of several years of having this measure having being reported.

In terms of the opportunity, I think it is 15 percent, and that is a subjective issue, I suppose, in terms of the number, but certainly in terms of the potential for improvement still

exists.

I will say that one of the issues -- I guess we can get that with reliability. There were some issues about potentially this proportion actually being underrepresented because of the measure being reflected on three days or less of antibiotic administration. There may be many situations where there are phone calls or follow-up for worsening of the illness beyond three days is not captured. So I think that is another example of where the opportunity is probably greater than what the number represents.

CHAIR SEPTIMUS: When we get into reliability and validity, we need to have a discussion about that.

DR. WINKLER: I have one question. This measure is related to children. Would it be appropriate for a measure similar to this for adults?

MEMBER FILE: You know, to me, it is sort of a corollary of what we just discussed.

Quite honestly, although acute bronchitis is technically a lower respiratory track infection, it actually accompanies most upper respiratory tract infections anyway. I mean, most people with common colds have a little bit of acute bronchitis.

So I think for the purposes to reduce overuse of antibiotics, I think it serves its purpose either way.

CHAIR BROTMAN: Yes, go ahead.

MEMBER HAVENS: But as has been pointed out, that gets to the point of the reliability and validity of the measures that we are talking about, and should focus our attention -- Since the performance gap here is the inverse of the performance gap in adults, then the question is: Is it because that, even though we all believe that we are measuring what we intend to measure, maybe we are not on the measure that we looked at before, or maybe this measure is something that we really wish we were measuring in the adult group.

So unless there are real data on reliability and validity that what we are measuring is what we want to measure, we need to be careful when we pass those criteria, because this is difficult to do with these level of data, and the exclusions here get you out a lot of different places that you don't get with the acute bronchitis exclusion in adults.

So the question that was just raised, should we be doing this in adults as well, should this be expanded to adults -- One question would be, if you applied this measure in a prospective study to compare using this criterion, comparing that to the acute bronchitis measure in adults, would the results be different? That would be one approach to getting further measures of reliability of validity potentially in that context.

So independent of what we do here, looking for the next time somebody goes to review the validity of these measures, we need to think

about why these are so different.

MR. HAMLIN: This measure does have a few additional competing diagnoses to it over the acute bronchitis measure in adults, to include common conditions in children such as pertussis and otitis media.

So I think that the specification itself would have to be revised before it was applied to an adult population to make sure that the appropriate competing diagnoses were included or not included, and those specifics would have to be tested to determine the effect on the rate overall.

MEMBER HAVENS: Absolutely, I agree with you, but that is the question about the reliability and validity of the prior measure and what might enhance the reliability or validity of this measure in this context. No, I appreciate your comments. Thank you.

CHAIR BROTMAN: I think those are valuable comments. Mohamad, please.

MEMBER FAKIH: Just to note that,

although the compliance with the right practice for kids is way better than for adults, if you look at this measure's performance over the years, it has not changed much.

So this does not mean that this measure is a better measure than the acute bronchitis measure. It just probably notes that maybe pediatricians who are not one of them do a better job than adults about physicians or it may be a different culture. You know, parents do not want advice. I wouldn't know, but just to show that this measure may not state what produced that result.

CHAIR BROTMAN: If there is no other discussion, let's vote on the performance gap at this point.

MS. KAHN: Voting on 1b, performance gap; again, it is 1 High, 2 Moderate; 3 Low; and 4 Insufficient evidence. You can start voting now. So we are looking for 20 responses. We are missing one person.

We have 3 High, 15 Moderate, 2 Low,

and zero Insufficient Evidence.

CHAIR BROTMAN: Let's move on to scientific acceptability with reliability.

MEMBER MURTHY: So again, similar to the prior measure, I think the reliability here, if we just sort of summarize again, their testing results reflect tow different sources that show very high reliability. I guess validity is a separate discussion, but I think, at least in terms of the testing approaches, the commercial and Medicaid report at the very end of the document supports a rate of .99 for commercial rate and one, very high.

CHAIR BROTMAN: Can I ask you, did it note the face validity?

MEMBER MURTHY: No. Sorry, under validity it was face validity only.

CHAIR BROTMAN: Any discussion? Let's go to the vote on reliability at this point.

MS. KAHN: Voting on 2a, Reliability: It is 1 High, 2 Moderate, 3 Low,

and 4 Insufficient Evidence. You can start voting now.

We have 5 High, 15 Moderate, zero Low, and Zero Insufficient Evidence.

CHAIR BROTMAN: Let's move on to validity.

MEMBER MURTHY: I think validity issues are similar to the prior measure where it was face validity and not the data elements, again very similar. It is a lot of description about the testing results, that essentially it is based on the face validity testing and not the data elements.

CHAIR BROTMAN: I have just a question with face validity. It was based on a panel. How large was the panel, and how extensive?

MEMBER MURTHY: It looks like the NCQA panel was made up of 21 members, reflecting sort of diverse members, and including quality improvement and scientific measurement, but I don't have more detail than that. Does that

answer the question?

CHAIR BROTMAN: Thank you. Any comments or discussion? All right. Let's move to the vote on validity.

MS. KAHN: Voting 2b, validity, 1 High, 2 Moderate, 3 Low, and 4 Insufficient Evidence. You can start voting now.

We have 2 High, 15 Moderate, and 2 Low, and 2 Insufficient Evidence.

CHAIR BROTMAN: Let's go on to usability at this point.

MEMBER MURTHY: I think the usability, again depending on the codes, the data sources, electronic records, again it is fairly usability in terms of the electronic health record program. Again, it is the same -- It is a little bit different from the bronchitis. I don't think there is a difference in coding anticipated. So potentially the usability for this in terms of tracking on a population level seem to be reasonable from reporting purposes. It is already part of the

CMS Physician Quality Reporting System.

CHAIR BROTMAN: Thank you. Any discussion?

MR. HAMLIN: I just wanted to comment. This is the one measure that did make it all the way through to a final published rule for use.

DR. WINKLER: Ben, this is Reva. You have already submitted the health plan level measure based on administrative data. What is your intention for that meaningful use measure?

MR. HAMLIN: As we get more reliability and validity testing accomplished for the meaningful use measure, we will be, certainly, providing that along with the specification itself. However, the rule was just published late last week, and we were waiting to determine which ones were finally published before we start including them in our applications.

Right now, only feasibility testing has been accomplished for the e-measures. So

we can probably be included that as well in our next update.

DR. WINKLER: Thank you.

CHAIR BROTMAN: If there is no other discussion, let's go to a vote on usability.

MS. KAHN: Voting on usability, 1 High, 2 Moderate, 3 Low, and 4 for Insufficient Information. You can start voting.

We have 10 High, 10 Moderate, zero Low, and zero Insufficient Information.

CHAIR BROTMAN: Looks like it is split between High and Moderate. All right, let's move to feasibility.

MEMBER MURTHY: I think we have already discussed some of this. It seems this is feasible in terms of the -- certainly, in terms of the data elements being accessible through electronic sources, and it sounds like, in terms of the strategy, it was already deemed appropriate for public reporting, and presumably would have had to have passed that bar for the meaningful use acceptance as well.

CHAIR BROTMAN: This is the one where it came up about delayed prescriptions. So anybody have any information on how to capture that?

MR. HAMLIN: The current measure, just for your information, does -- is off of dispensed prescriptions because of the administrative claims nature of the measure. We do have the option in the EHR measure to look at prescribed versus dispensed, and we have additional options for future measures in the future to determine the time frames between those two events as they occur. However, we are limited to the administrative claims dispensed information.

MEMBER MURTHY: So is there an opportunity then to extend the time from three days to further out? Is that what you are indicating, in terms of dispensed?

MR. HAMLIN: I think that is one of the considerations that our panels will be looking at for the future measure to determine

what is the appropriate time frame.

The window was determined during field testing originally when the measure was first published, that three days was the appropriate time frame due to the other comorbid conditions and the appropriateness for the antibiotics in this population group, but I do expect we will be looking in the future at not only the time frames, but also the different types of encounters that could potentially occur and how those are being administered to the patients.

CHAIR BROTMAN: If there is no other discussion -- Tom, I'm sorry.

MEMBER FILE: Along those lines, I recall in the discussion in the Cochrane Systematic Review for acute bronchitis, they had a significant discussion on strategies for reducing antibiotics, one of which is sort of delayed prescription type issues, and they presented, at least cited, various studies that did this where there was about a 50 percent

reduction.

Interestingly, however, because this may be something that needs to be considered in meaningful use and consideration, there was a significant decrease in patient satisfaction with that. So if you are going to measure a patient's satisfaction as part of quality of care as well, you have to take that into account.

MR. HAMLIn: Yes, and as a matter of fact, in the adult measure, if a provider does write a prescription for the patient at the encounter, however, gives specific instructions not to fill it unless symptoms persist for the next seven or eight days, they will actually be numerator compliant if they do not dispense it and vacation within the seven-day time window. However, the evidence or the results would indicate that patients are probably not complying with that seven-day window and just going and getting those prescriptions filled.

CHAIR SEPTIMUS: I have a stupid

question. with electronic prescribing, tell me mechanically how this works.

MR. HAMLIN: In the e-measure, we look at the dispensing date, as we do with the administrative claims measure, but again I expect to look at in the future measure when the prescription was entered into the system, when the medication was, in fact, dispensed, and we will have very accurate information to the minute in some cases as to that time window, which will result in a -- which will necessitate, I should say, additional discussions about what the appropriate time windows are.

CHAIR SEPTIMUS: So if I do an electronic prescription for an antibiotic and I tell the patient to wait three days, is the pharmacy not going to fill that prescription and wait for the patient to pick it up? That is what I am confused about.

MR. HAMLIN: I would certainly hope not, but again this is where I think we have to look at in the e-measure.

CHAIR SEPTIMUS: How would the pharmacy know that?

MEMBER ELAM: I can speak to that from Kaiser's standpoint. We are able to do prescriptions and put them on file, and so they are not dispensed, but they are on file so, if the patient activates that prescription three or four days down the road.

MR. HAMLIN: So pharmacy is one of the things that is getting a lot of scrutiny in the e-measure world, and we are certainly very aware of the differences in practices across different platforms.

MEMBER GIORDANO: I just wanted to see if anyone had any other thoughts on that, because that is great that Kaiser does it. I am just wondering if others do. That doesn't sound pretty standard in the electronic pharmacy prescriptions that I know.

CHAIR BROTMAN: Kathleen?

MEMBER BRADY: No. I would say that is not. I have never heard of that before,

actually.

CHAIR BROTMAN: All right. If no other discussion at this point, let's vote on the feasibility aspect.

MS. KAHN: Voting on feasibility, again it is 1 High, 2 Moderate, 3 Low, and 4 Insufficient Information. You can start voting.

We have 4 High, 14 Moderate, 2 Low, and zero Insufficient Information.

CHAIR BROTMAN: And let's just go on to suitability for endorsement at this point. Is there any further discussion? Anyone have any comments? If not, we will go to the vote.

MS. KAHN: Overall suitability for endorsement: Does the measure meet NQF criteria for endorsement? Vote 1 for Yes, 2 for No. You can start.

We have 20 Yes and zero No, and the measure will pass.

CHAIR SEPTIMUS: Well, I want to thank the developer and the presenters. We are

due for a break. I think we are going to try to make it a 10-minute, unfortunately, though, not the 15 minutes. The restrooms are outside this door, past the elevators, and I think you turn right. So we will see you back here at about seven or eight after eleven.

(Whereupon, the above-entitled matter went off the record at 10:57 a.m. and resumed at 11:11 a.m.

CHAIR SEPTIMUS: Okay. The next measure is 0500, Severe Sepsis and Septic Shock Management Bundle. The developer is Henry Ford, and here is Dr. Rivers to give a quick oversight, and then Tiffany will go through the measure. Dr. Rivers?

MR. RIVERS: Thank you so much for allowing me to come. I actually was at the last NQF meeting -- I think it was three or four years ago -- presenting the measure primarily, and it was essentially endorsed at that time. So we are here today to talk about a revision as well as the maintenance aspect of the measure.

What this is all about essentially is what I will call common sense practice. I both work in the emergency department and critical unit at Henry Ford, and been there for almost 25 years, and one of the things I noticed about a patient who could come in infected is that they will lay in the ER for 12-14 hours in septic shock, and by the time they got to the ICU there was nothing you can do for them. So essentially they died. The mortality was over 50 percent.

Now this is no small hospital. this is a hospital that won a Malcolm Baldridge Award this year, and in 1997 we had a septic shock mortality of over 55 percent. So you could literally walk in there and get a liver transplant. You can get a kidney transplant or heart transplant, but you would die from sepsis.

So that paradigm, we couldn't tolerate. So we started a quality initiative, not a study but a quality initiative. So what

we did is search for what we call a standard operating procedures for sepsis, and we looked around and found the Society of Critical Care Medicine and American College of Critical Care Medicine had some protocols, and we started to simulate these protocols, along with expert opinions as far as 1997 by Robert Wilson from Wayne State University.

This comprised what we call a sepsis operating procedure. So we did a study in which we actually had to randomize patients, although the control group was not truly a control group. We saw a mortality reduction of over 16 percent, and we instituted this as a standard of practice.

So from the year 2001 to 2007 we accumulated over 2000 patients, and we showed a mortality reduction from over 50 percent down to less than 10 percent, which we actually have today. So this is what we call the Henry Ford measure, and we presented this to the NQF back in 2007 and started it in 2008.

Since that time, the measure has been implemented amongst over 54 publications, numbering 20,000 patients over the last decade. Mortality reduction in patients of equal illness severity over 40-45 percent have been 14 to 16 percent consistently, with an average reduction in hospital stay of about five days.

So what have we done since? Well, the key point is operationalization. It is a protocol telling about a patient that basically is owned by no specialty, and so, therefore, emergency medicine, critical care medicine, etcetera, has had issues in terms of who owns this patient from a hospital perspective.

So one of the great challenges is basically to create a hospital-wide initiative versus a specialty related initiative, and in doing that we have a combination of emergency physicians, critical care physicians. We have floor physicians, floor doctors, even what we call mid-level providers, who all get together, and we manage this septic patient as a

hospital-wide initiative, not a specialty related initiative.

So what we have seen over the last decade is programs like Kaiser Health Care, Catholic Health Care West, Capital Health Partners, Intermountain Health, HCA Healthcare, comprising over hundreds of hospitals that have seen the same mortality reduction we have.

So here we are today revisiting NQF to reinforce this as a true measure that can be extrapolated to better patient outcomes.

So what we liken this disease, I think, in summary, is a heart attack where you came in with a heart attack 30 years, they gave you some oxygen and then aspirin. Now you have thrombolytics. Now they take you to the cath lab with a door-to-needle time of 90 minutes, and this is actually a quality measure, meaning that if you don't meet these criteria, there are incentivized as well as the incentivized ramifications. So it is simply an evolution.

If we look at trauma patients and

we also look at stroke, the same evolution. So this is not a novel concept. What we have is an evolution of a disease that, number one, needs to be treated very aggressively, which accounts for over $60 billion in Medicare related costs, the most expensive hospitalization in the United States since 1997, and carries the highest mortality, almost nine times any admission to the hospital. You are more likely to die nine times greater from sepsis than any other disease.

So with that, I bring to you this measure, and I appreciate this opportunity and will be happy to take any questions. Thank you.

CHAIR SEPTIMUS: Thank you, Dr. Rivers. With that, we are going to turn it over to Tiffany, who reviewed this measure for her Work Group, starting off with the impact.

MEMBER OSBORN: If it is okay, just prior I would like to make a couple of comments, if that is okay. Okay, great.

CHAIR SEPTIMUS: I could never say

no to Tiffany.

MEMBER OSBORN: That is really appreciated. I just want to make sure for people who have called in on the phone that, for the sake of transparency, that my disclosures again have been made, that I have been a Sep Representative to the Surviving Sepsis Campaign for over a decade.

I have assisted with the Institute of Health Care Improvement in implementing locally determined versions of really goal directed therapy into a health system. Then I have also served and currently still serving as the trial clinician for a study called ProMISe, which is Protocolised Management in Sepsis, which is evaluating early goal directed therapy within the context of the UK system. So I want to make sure that that is clear.

Additionally, I think that the committee should know I have received numerous -- and I do mean numerous -- communications regarding this measure, and I think it is

important that I allow you that I provide as objective information as I can, but there is more than one way in which this data has been interpreted, and I should probably provide that.

Dr. Rivers has talked about the fact that there is a lot of studies currently that have been done on this topic. There are almost 60 studies, maybe a bit more, encompassing 50-60,000 patients, the majority, the vast majority of which has demonstrated survival benefit.

To my knowledge, no study to date has demonstrated increased mortality. This meta-analyses, all meta-analyses that I have seen up to this point have shown survival benefit, and it has been the premise upon which both national and international guidelines have been created on the management of severe sepsis and septic shock. However, there is also an alternate view of that data, and that is that -- Well, prior, let me just say that as a result of that information, there are a number of people

and groups that would advocate that there is enough data, enough data exists to implement CMS measures and that potentially waiting or delaying this could potentially risk lives for very little gain. So that would be the way one contingent would see that.

A second contingent would see this as the vast majority of studies save one or two, were observational. They were bundled completion/incompletion studies. They were before/after studies, and they would also submit that these are subjective to inherent bias.

Additionally, there are three ongoing international trials that are evaluating this bundle, and there are plans to do a patient level meta-analysis at the end, and some would advocate that these trials would present valuable information to the discussion.

Now there is another group, another contingent, and I would think that the American College of Emergency Physicians would fall into this. There is a group that believes that the

information and the data that currently exists is valid, but they have questions regarding the implementation.

Their thought would be CMS having grand rounds in march, if this were reviewed and voted upon in March, that it would still be included in the inpatient and outpatient proposed CMS rules of 2013 and would not delay measure implementation.

Additionally, I think it is important to review that there is one component of the measure, one element of the measure, that I received a number of emails about, and that was regarding central venous pressure.

So in the management of acutely ill and injured patients, many of us use central venous pressure as a surrogate measurement to estimate intravascular volume, and there is a contingent of people who think that the use of CVP, that there are a number of studies that would say that it is an inaccurate measurement and that they feel -- this same contingent would

feel that clinicians would be limited if they were -- because there are multiple ways to measure intravascular volume, some of which some people feel may be able to measure intravascular volume more effectively than central venous pressure. They don't feel that they should be penalized for using something else.

CHAIR SEPTIMUS: Tiffany, I hate to interrupt you, and those are great comments, but why don't we go ahead and go through the measures and, as we come to sections that apply to your comments, we can have some discussion, but let's go ahead and start with the impact, and let's go through the same list that we went through with the previous measures, and then those other comments can come in as appropriate for those sections when we start talking about validity, usability, etcetera, if that is okay with you.

MEMBER OSBORN: Okay. So the first on importance to measure and report: As presented previously, there are a number of

studies that are currently out that show survival benefit and, as stated previously, none at this point, to my knowledge, have demonstrated harm, and this has been used for both meta-analyses, national and international guidelines, and as stated previously, there are other ongoing randomized controlled trials that are currently pending.

CHAIR SEPTIMUS: Any discussion now about the impact? Peter?

MEMBER HAVENS: The impact is enormous, clearly. The question of the central venous line is a critical question. If you are faced with a patient who can't get a central line or has severe sepsis with DIC, then that patient is excluded from this, as I understand the denominator exclusions. Is that accurate?

CHAIR SEPTIMUS: Peter, can we hold that to the scientific validity. We are only talking about impact.

MEMBER OSBORN: I would just add on impact that right now there are greater than

750,000 estimated cases of severe sepsis a year in the United States. Additionally, there are an estimated 400,000 ICU admissions, around 200,000 deaths a year, and it costs an estimated $17 billion a year.

MEMBER HAVENS: I apologize for bringing up that question at the wrong time.

CHAIR SEPTIMUS: Don't apologize. I am just trying to keep people on track, because we really want to adhere to the exactly the same format. So we are only voting -- will vote first on the impact, and then we are going to get to evidence and opportunity, but let's stick with the impact first.

Any other discussion about impact? then we will vote.

MS. KAHN: We are voting on 1a, high impact, again High, Moderate, Low or Insufficient Evidence. You can start now.

That is 19 High, one Moderate, zero Low and zero Insufficient Evidence.

CHAIR SEPTIMUS: Thank you. Now we

are going to go to the evidence.

MEMBER OSBORN: I think that I have described the evidence pretty clearly already, and if there are any questions --

Dr. Rivers.

DR. RIVERS: I just wanted to mention, there is also the evidence of outcome benefit, and there is also evidence of the individual bundle elements. So if you take the studies and you do what they call regression analysis of each bundle element, there are studies that support each element within these studies.

So if you isolate CVP, there are studies to show that CVP actually relates to outcome, which has always been sort of a point of contention, as Tiffany said, but I think it is very important that these studies that have been looked at actually show that CVP is an impactful endpoint for outcome. So is SCV-02. So is mean arterial pressure. So is antibiotics, and so are other of the variables

that are within this protocol.

So even within the sub-analysis, there is outcome benefit. Although they are discussed as controversial, they are within those studies.

MEMBER OSBORN: With regard to CVP, additionally, I stated what the group who would advocate against it. I gave you that information. The group that would advocate for it would also say that, in general, central lines are -- If you have a patient who presents in septic shock, that patient requires a central line for vasopressor use, and that measure of CVP is a natural extension of that.

Additionally, that group would state that the trend -- the importance would be the context of the trend in context with clinical symptomatology and that that is easy to follow at the bedside looking at the monitor.

Additionally, as Dr. Rivers pointed out, there are a number of studies that require the estimated intervascular volume as part of

their study, including vasopressor studies, recent ones that have been done within the last few years that used CVP.

Finally, one person brought up something that was actually quite helpful in stating that measuring CVP does not actually preclude the use of any other method to measure intervascular volume. It only states specifically what you will be measured on as far as the quality component.

CHAIR SEPTIMUS: Tom?

MEMBER GIORDANO: Could someone clarify if we are supposed to be evaluating the evidence related to whether CVP is important or whether measuring CVP is important. Obviously, if your CVP is too low, you are dead, but is measuring it important? Is that what we are supposed to be evaluating, and the same for all -- I mean, the same with lactate, for that matter.

DR. WINKLER: Right. Essentially, the evidence is to look at that process of care.

What do we know from studies that that process of care is related to patient outcomes. So if the measure is about measuring it, then that is the process of care. If, for instance, the measure were about a specific level, then that would be what you were looking at.

So the evidence is exactly what the measure is constructed. You are asking, do we have -- you know, what is the scientific basis, the literature behind that process of care as defined in the measure.

CHAIR SEPTIMUS: Mohamad?

MEMBER FAKIH: Just a clarification and a question. You know, when we look at evidence, what my worry is -- Now this is a great protocol to do for severe sepsis and septic shock, but is it the best protocol, because when we look at this to become a measure, it is going to trump all other ways to do the work, and that is my worry.

Is this where we are going to look at this, as evidence? So the best evidence --

you know, the best approach -- We are going to look at it as a good -- There is evidence that it works, but the question for me when I vote on this: Is this the best approach, because it is going to be very hard to have it as a measure, and then all other competing protocols would be trumped, because you have to follow this measure.

I may be going too far, but --

CHAIR BROTMAN: Just to remind you, we are looking at the quality, consistency, and quantity of the evidence presented.

CHAIR SEPTIMUS: Again, we went through this, but just to remind you, the quantity talks -- you know, for High it is five or more studies. Moderate, it is two to four studies. In terms of the quality, we looked at randomized controlled trials as being the highest. Moderate is nonrandomized controlled, but it could be a large study with a large impact. Low would be ones that are significantly flawed and introduced bias. So

just to give you that rundown in terms of the quality of the evidence.

Then the consistency -- Oh, we got it up there. I think you need to look at this, and the consistency has to do with stability in both direction and magnitude of the clinical and practical, meaningful benefits.

High would be that it benefits and little harm. Moderate would be at least one study that estimates the benefits greatly outweighs the harm. Then low, of course, there really aren't very many good studies.

So that is what we are looking at. Then the composite is shown here on this slide.

MEMBER GIORDANO: So are there randomized data on the bundle, whether if the process is -- if the bundle as a process is implemented compared to where the bundle was not -- where there was no monitoring of the process for the bundle? Does that make sense? That is what we are being asked to evaluate, right? Is the bundle study in a randomized

trial -- what is the quality of the evidence for the bundle as a process of care?

CHAIR SEPTIMUS: Dr. Rivers, did you want to -- The developer can respond to these individual questions. Did you want to respond to Tom's comment about the bundle?

DR. RIVERS: Sure. A couple of comments in reference to evolving what we call standards of care. We commonly have that today. If you look at acute myocardial infarction, it is looked at every two to three years, and all the components, whether controversial or not, are revised.

Same way with advanced cardiac life support. There are parts of advanced cardiac life support that have probably one trial. That is not even randomized, but it is put in there as the best evidence to date, and those things are in evolution.

With sepsis, there is no standard. For the first time, we have a standard for a disease that kills almost half the patients that

get it, but there is no standard. So I think it is important to understand that this is an evolving process, and what we don't want is to say we will lock in something that won't -- obviously, can't change.

I can tell you, in 2008 when I presented this to this committee, they were proposing the same trials that are going on today, as though these trials were going to answer the questions. Four years later, these trials have provided no information. So we are four years later waiting for some clinical trials to get finished, but people are dying, in essence.

A recent trial just published last week is called Genesis, and this took place in 11 hospitals throughout the U.S., and these hospitals range from 100 patients to 1,000. They took this protocol. They said, implement it. What they showed is the 14 percent mortality reduction.

Now, granted, it was a before and

after cohort, but what was unique: There was a prospective cohort, a bundle met, bundle not met. So in that subset of 6,000 patients, there were 1,000 patients where they compared people who met this bundle and did not meet that bundle. The mortality reduction went from 44 to 30 percent.

So just by meeting a bundle in a prospective observational cohort -- Now people say, oh, this is not a prospective randomized trial. Well, this is a prospective observational not only in large hospitals but small hospitals as well that show the mortality reduction. This was just published in the Journal of Critical Care just last month.

CHAIR SEPTIMUS: Okay. We've got one, two, and three. So, Aaron.

MEMBER MILSTONE: I wanted to follow up on a similar question. I was looking at the actual proposed measure where it talks about the overall bundle compliance, and I wanted to get more at this what I think is a

fundamental question, which is: We are measuring the evidence of the bundle, not of each individual measure, and there are a couple -- a handful of references here, but I wonder if you would just expand a little more on how your group interpreted these couple of studies that look at a bundle, and were all these this bundle or were they just any sepsis bundle?

The study that you are mentioning that just came out I didn't see in here. Was that this proposed bundle?

DR. RIVERS: Exactly the same.

MEMBER OSBORN: It is a very good question, and Thomas' question still remains to be answered, and his question was how many randomized controlled trials are there. Perhaps Dr. Rivers can answer that.

My understanding of this bundle, that there is one currently, maybe two. Can you --

DR. RIVERS: There is the original one in 2001. There is one in China. There is

actually one in Taiwan, and there's two eastern Asian studies that just came out, and actually Brian Wynne was the investigator, and those were what they call prospective trials. But the key point in terms of randomization: There is the issue of equipoise, and equipoise means that can you legitimately allow a patient to have a control group which you will do nothing or basically wild type standard of care versus what we know as best practice.

So when we talk about randomized trials, what we have to understand is that you are subjecting people to a basic standard of care over what we know best since 1967. That is expert opinion.

MEMBER OSBORN: The other question that was asked here, though: Were they the same bundles? If I remember correctly -- you can help me, please -- that specifically the one in China used CVP but not SCV-02. So not all of these trials actually had all components. Is that correct?

DR. RIVERS: What we have to understand, these are quantitative resuscitations, but the majority of studies have used this complete bundle, and the study I referred to called Genesis has used exactly all elements of this bundle. That was 6,000 patients.

MEMBER OSBORN: And that was a randomized controlled trial and an observational trial.

DR. RIVERS: It was observational cohort and a prospective observational, because they did not want to randomize patients to what we call standards of care.

MEMBER MILSTONE: There is one observational -- and not that this is bad, but there is one -- Just so we know the data, there is one observational study looking at this complete bundle?

DR. RIVERS: No, there are 55 studies out there, and I would say, if you look at variations of what we call "this bundle,"

this exact bundle, 40 out of 54 patients -- Forty out of 50 of this studies are what we call identical to this bundle.

CHAIR SEPTIMUS: Kathleen?

MEMBER BRADY: Yes. So that is -- My issue is that with the information that was submitted, it is really unclear to get an idea of how many RCTs, how many observational studies, how many of the studies use the exact same information. There is just nothing in this that really gives me an overall summary of how many patients that involved, etcetera.

DR. RIVERS: Well, if you want to understand --

MEMBER BRADY: It is just -- maybe a list. That is my issue. Is there a list of 55 studies that I didn't have time to read?

DR. RIVERS: Right. I understand. Well, if you look at 50 -- at least 40 of those studies are basically identical in terms of their protocols. Some people have variations in terms of -- like you say, some people may

not have looked at SCV-02, but they completed the whole bundle aspects.

So the key point is to understand that in those studies, at least 40 to 45 of those studies are basically identical in terms of this management, and that comprises over 20,000 patients.

MEMBER HAVENS: So did -- The population of study in those protocols presumably would include people who did not have CVP monitored? The reason I ask is trying to understand the population of measure here, which excludes people without CVP monitored.

So if the goal of this outcome measure is to look at all people with sepsis, which is defined here, but excludes people who don't get a CVP for a variety of different reasons, which may be practice related or disease related, then there is a terrible do loop of inefficiency in this measure as constructed. Do you see my problem here?

CHAIR SEPTIMUS: This is a process

measure.

MEMBER HAVENS: No. No, this is a fundamental question about what we are actually trying to study. Are we trying to study people with sepsis, and the studies you are saying that apply this model show benefit only if they get a CVP then? That would be the benefit included, because you couldn't study it in another group.

DR. RIVERS: Well, I understand what you are saying. When this study started, nobody looked at one element. It was like driving a car. You have the brakes. You have your lights. You have your accelerator. You drive a car to the store and back. It is not like you look at each -- So this was never the intent even back when the study started.

As we evolved, people wanted to come up with a less complex way of managing these patients. So that is why individual bundle elements. There are six or seven other elements that we are not even talking about. Do you need to give the patient antibiotics in three hours?

Do you need to give the patient fluids?

So I think the fixation on CVP has to come about, because it is the most difficult, and it requires a technical expertise that requires -- some physicians in emergency departments cannot perform.

So what we have to understand is that there is a technical barrier here that requires a procedure that is more expert, and that is the difference. You have these patients who are just as sick as ICU patients in a place where they perhaps do not have the level of competency to manage those patients, and they should not be there, in essence.

That is the essence of this whole issue. It is not -- When those patients go to an ICU, they get a central line placed. They just happen to be in a ED, and that is the nature of care in this country.

CHAIR SEPTIMUS: Okay, next question.

MEMBER CAMPOS-OUTCALT: I have one

question and one comment. First of all, the question for the staff: How common are these types of bundled measures?

DR. WINKLER: Fairly common. We are seeing more and more that are sort of all-or-none composite measure, if you will. You must complete all elements of it to get credit for the whole measure. So, actually, they are common and growing.

MEMBER CAMPOS-OUTCALT: Okay. So a couple of comments. I have some sympathy with the view that was expressed earlier regarding once you set the standard, you can't study afterward. There are a number of variables here that -- I am sure that antibiotics within three hours is better than not, but is four hours as good? Is five? And we are never going to be able to study those things after this. So that is one thing that is bothering me.

Secondly, we are hearing some comments regarding evidence which, sitting here trying to make a decision regarding evidence,

I find somewhat unsatisfactory.

Observational studies by themselves are not necessarily poor evidence. If you've got a number of them, they are consistent, they have high observationals and so forth, you can upgrade them to high quality evidence.

I haven't heard anything about an independent evidence report that does that, and I hear an issue where there is controversy between two factions, and I am only hearing one side. I would very much like to hear the interpretation of the evidence by the other side, in light of the fact we don't have an independent evidence report.

So I have to say that at the moment, I am standing here saying this is insufficient evidence for me to decide.

CHAIR SEPTIMUS: I am getting some suggestions that I think are good ones. First of all, there are actually -- In terms of delay in antibiotics by hour, there actually is data for that. So I think perhaps what we might do,

since we are getting hung up between individual elements of a bundle and the bundle as whole, maybe as our very wise people here to my right indicate, then maybe we should go through each of the bundle elements individually and look at the quality of the evidence.

We use bundles all the time in HAI prevention. That is the standard. We don't just do one thing. We do them all. So bundles in health care are very, very common for those of you, but if it would be helpful, we can go through each bundle element and look at the quality of the evidence, if you would like, if that will help.

MS. BOSSLEY: Hi, I am Heidi Bossley. I am with the NQF staff here. I was actually next door in the GI/GU meeting, and they just went through a similar measure that had multiple components, and it was very helpful to walk through, first of all -- and there's two issues.

Number one, were you provided the

information? Then, if not, is there information that people collectively at the table are aware of, evidence, related to evidence for each of the individual components, and summarize that information.

Then you can have -- Again, you have the three options to vote. If there is information that you are aware of that is not in the form, you can vote it down, no, insufficient information provided, and then we can move on to the next slide, and then you can have a discussion on, if it was provided, how did all of these components in this bundle rate on the evidence?

So I think it would be very helpful. It would be very transparent when this goes out for comment as well, so that people understand how this measure was voted upon, to go through each explicitly. Does that make sense to everyone?

CHAIR SEPTIMUS: What is the consensus of the group? We still have some

questions. I haven't forgotten you, but there are still some questions outstanding.

MEMBER FAKIH: My concern is that, when you look at the individual points or individual parts of the bundle, having an all or none does not mean -- So adding them altogether present does not mean that they give you the same result, having all or none versus having four out of five.

You know, we are assuming that, if you have the five, let's say, points together, this is better than having four out of five. I don't know if that is the answer, and that is really tough for me to vote on, you know, saying it is an all or none bundle.

I think we ought to take it as individual, and we push for it as individual part, measures, or we can't do it otherwise. I can't --

MEMBER OSBORN: May I present some information that might be helpful on that?

CHAIR SEPTIMUS: Aaron, you put

yours down? Okay. Adam?

MEMBER THOMPSON: So this is a question to the people who looked at this more in depth. One of the options we have on here, too, when the empirical evidence is being questioned, is does this process of care bring greater good than bad?

I am just questioning, for those people who are part of this Work Group, is that something you would consider making an exception for, having read through this more thoroughly? Does this benefit us as patients more than it would hurt us, in the absence of some other process?

CHAIR SEPTIMUS: Go ahead.

MEMBER CAMPOS-OUTCALT: My comment is going to be that, if we are going to have the same process, which means there may be some controversy over the interpretation of the evidence and we are only going to have one side, I won't find going through these individually that helpful.

CHAIR SEPTIMUS: Michael?

MEMBER FARBER: I have the same issue, and that is that the way I originally interpreted any of these questions is that the measure as a total is all the bundles. So that is what I felt we would be voting on. We still could do that, but the issue is: Let's say, if the bundle includes five, six, seven components, if two of the components really have not been demonstrated to be worthwhile, we are now including these as part of the measure.

So I think this is a problematic measure, but I still think it could be voted on, if one wants to determine to vote it on as a total. But I think it would be a big problem to break it apart now, because then we would be breaking apart to five or six separate measures, and the sum of the parts may not equal the total.

So I would vote either to continue to vote on it as it was original, as a bundle, or to drop it altogether.

MS. BOSSLEY: Can I just clarify? Perhaps it wasn't clear what I was suggesting. I think you need to have a discussion of each of the individual components that make up this all-or-none. At the end of the day when you go to vote on evidence, though, it will be whether those components altogether pass the evidence. So is there consistency, quantity, and quality that is needed at the ratings that we had, we provided to you, for all of those components?

I think, again, I am hearing some -- I am not sure, if you voted on this now, if I could understand whether you were voting it because there was issue with one component within this because of the evidence or if there is actual multiple. That is why I think it would be helpful to walk through the evidence for each of these individual components.

CHAIR SEPTIMUS: And just let me clarify, and tell me if I am wrong. The measures are out there, but you don't have to do them

all. A lot depends upon how the patient is. So if you look at the measures, and tell me if I am wrong, there is lactate, blood cultures, antibiotics, and fluid resuscitation for hypotension or lactate greater than four.

Then, depending upon how the patient does, you will then apply vasopressor for patients who remain hypotensive despite fluid resuscitation, and then if you continue hypotensive, meaning you are septic shock, and you have a lactate greater than four which would indicate micro-circulatory issues, then it would be worthwhile to measure CVP and SCV-2. So in other words, it depends upon the patient, how the patient does. It is not all or none. It is we do these and, depending upon the response, we may do additional things. Did I get that right?

MEMBER OSBORN: I think most data actually shows that only about 15 percent of patients get down to requiring SCV-02, if the other components of the bundle are followed.

Additionally, this discussion about looking at each individual bundle component has come up before, and probably people from the Surviving Sepsis Campaign that are on the phone could answer to this better than -- you know, would be better people to answer to this, but this topic has come up before, and the data that we have, really, about how this impacts mortality has to do with implementation of the bundle as a whole.

CHAIR SEPTIMUS: Dr. Rivers.

DR. RIVERS: Yes. I have actually looked at individual bundle elements and have that data. So whether you want it now, I can provide that. There are literature that have looked at each one of these bundle elements and given regression equations and various aspects to see whether each bundle element has an impact on outcome, and that data does exist.

Now when you do that kind of data analysis, it is usually based on examination of a cohort, and they do multiple regressions,

balancing for illness severity, to see if they can isolate that bundle element. So those are now prospective data that you can look at.

If you want to say -- each one of these bundle elements, and say we got six or seven. We multiple that times 1,000 patients, we will need a 6,000 to 10,000 patient study in order to come up with whether or not CVP, mean pressure, or all these elements actually impact outcome.

So it is a very complex question, but if you look at the literature and say, well, what does the literature say in those observational studies, which elements are important, there are a number of articles that have looked at those, and they remain statistically significant with mortality.

CHAIR SEPTIMUS: Michael, did you have another question or you just hadn't put your -- Okay. Sir, did you have another comment? Okay. So let's see if we can summarize. We can either go through each

element, which I am gathering people are not terribly enthusiastic about, or based on what we have already heard, we are ready to vote on whether or not there is sufficient evidence, because this is one of these stop things for the measure.

MEMBER OSBORN: I think that there might be another piece of information, and again I am trying to objectively provide both sides of the equation.

My understand -- and look at the last meta-analyses that were done on this of the various studies that were listed, I think maybe 15 of those actually contained -- and that is an estimate; you know, I would have to go back and look at all the new data. I am talking about the meta-analyses that looked at the bundles that actually examined each -- that compared or put together the bundles that had the exact same components, meaning did they give fluid, did they measure CVP, did they apply vasopressors, and did they correct SCV-02.

My understanding is there's probably about 15 or so trials that had all of those bundles, based on the meta-analyses that were published. I know that there are some others that have come out since.

DR. RIVERS: That was by Chamberlain out of Australia and New Zealand, and that was over 12,000 patients, and it actually isolated each bundle element. But if you look at the growth of publications, there is now 54 to 55, and nobody has done a recent meta-analysis comprising those studies. So you probably have one that represents about half the studies out there.

MEMBER OSBORN: So you think about half actually use all of the elements of the bundle, including SCV-02?

DR. RIVERS: And Chamberlain's study -- Chamberlain did that.

MEMBER OSBORN: Great.

CHAIR SEPTIMUS: Not to get too deep in the weeds, but Mitch Levy in Critical Care

Medicine published the Surviving Sepsis bundle, and just to -- Again, tell me if I got this wrong. they were looking at both management, which is not going to be in the upcoming version, but also resuscitation, and with increasing compliance.

This is an all-or-none. You didn't get credit unless you got it all done. What they found was with increased compliance, they saw a statistical reduction in mortality. So there are lots of observational articles that, I think, match what everybody else is trying to say.

DR. RIVERS: I think it is important to realize that your compliance in real life of 60 to 70 percent is what the national average is. It is not 100 percent. You are never going to get 100 percent. So like you say, you get a patient who comes in with DIC that has coagulopathy, you can't put a central line in them. That is the reality of practice. Some patients come in so far along that that is not

possible.

So what you do is you have what we call compliance ratios and compliance ceilings and, actually, about 50 to 60 percent you will see an incremental decrease in mortality, not 80 to 90 percent.

Brian Wynne published in 2007 where the cutoff in terms of compliance improvement and mortality. It actually occurred around 58 percent. So if you just had a 58 percent score, you saw a reduction in mortality of over 15 to 20 percent.

CHAIR SEPTIMUS: Okay. Any other -- Yes?

MEMBER CAMPOS-OUTCALT: So if the evidence is so clear, help me understand what the controversy is. Why do we have a group out there that opposes? You have been very fair in presenting, and I appreciate that. So maybe you could help me understand. If the evidence is so clear, why do we have such disagreement?

CHAIR SEPTIMUS: Okay, let's go to

Aaron, Mary, and then Tom.

MEMBER MILSTONE: I actually have the same question. You initially presented this controversy about patients, measuring CVP. Just so I am clear, if we think the evidence is sound, why there is this controversy of two camps, because again this would create a standard of care that likely will be -- It is easy to make bundles, but it is hard to get elements out of bundles after the fact.

MEMBER OSBORN: I really wish that there were people who are on the other side of the camp who were here to talk so I wouldn't have to be the one to talk for them. However, there are really three. There are two major categories, and then there is a subdivision of one.

So there is the camp, as I stated before, that would say, look, you know, this data is valid. There are a number of observational trials, and that camp says this data is valid, and we should move forward with

this data. And they would say, there is close to 60,000 patients right now who have been evaluated.

Then of the other camp, there are two subdivisions. There is the group that says, this is best practice and shouldn't be implemented as standard of care. There is a contingent that says that, and there what they put behind that is the fact that all of these, as stated before, except for may two are observational trials and that it is inherent to bias.

Then there is a third camp that says, look, we believe it is valid, but we think that the actual implementation of this still needs to be worked out, that how it is actually put into practice needs to be specified more.

So it really comes down to how you interpret the data whether or not you are someone who believes, look, we have a number of observational trials. There is a group of people that feel that doing randomized

controlled trials in this patient population may be unethical.

There is clearly another contingent. There is, like I said, three ongoing randomized controlled trials, one in the U.S., one in the UK, and one in Australia, that are evaluating this very same measure and are going to do a patient level meta-analysis at the end.

So there is in some areas equipoise. So it really comes down to how you value the level of data that is currently available.

CHAIR SEPTIMUS: Okay, Kathleen. Hey, she defers to you, Tom. That's great. I am sorry if I did not get the hands up. Tom?

MEMBER GIORDANO: So let me try to clarify this then. There are meta-analyses to date, and the observational data are on the process, and they show that the process matters, not that lactate predicts survival but that measuring the lactate sooner predicts survival. Is that correct?

MEMBER OSBORN: There are a number of studies that demonstrate that normalization of lactate does impact survival.

MEMBER GIORDANO: But that is different. Measuring lactate sooner, that measuring CVP sooner, that instituting antibiotics sooner impacts survival. Is there data for that?

MEMBER OSBORN: Yes, there is data for that.

MEMBER GIORDANO: Then the randomized controlled trials are, again, randomizing to process. So, yes, you either get the bundle or you get standard of care, which might include some elements of the bundle but might not.

MEMBER OSBORN: The randomized controlled trials that I know of to date are the sentinel study that was done in 2001 by Dr. Rivers, which measured doing all the components, meaning fluids, measuring CVP, then instituting blood pressures to a certain mean arterial

pressure, and then normalizing SCV-02 as indicated, versus giving fluids, measuring CVP and using urine output. That one demonstrated a 16 percent mortality benefit.

The other randomized controlled trial that I know of, and perhaps Dr. Rivers can assist if there are others -- The other randomized controlled trial that I know of, the other ones, look at those components up to CVP and don't include SCV-02.

MEMBER GIORDANO: But you mentioned that there are some ongoing randomized trials.

MEMBER OSBORN: Yes, sir.

MEMBER GIORDANO: So there must be enough uncertainty out there among the experts, among the funders, to question to sponsor major, very expensive randomized trials. Is that a fair assessment, and are those randomized trials of process?

MEMBER OSBORN: They are currently ongoing. Probably -- Two of them are probably -- They are probably all about the same amount

through, probably about halfway through, a quarter to halfway through.

So to get back to your question, it really depends on how you evaluate the data that is in front of you. If you look at the quantity and the quality of observational trials, including the Surviving Sepsis one which was, in a sense, bundle completion and completion study.

If you look at those trials and you think, look, there is enough quantity and quality of observational trials that I feel comfortable, then that is one thing. If you are of the sentiment that I need a randomized controlled trial, then it wouldn't meet your threshold. So that is sort of a -- That is an individual practitioner threshold, I think.

CHAIR SEPTIMUS: Just to follow up on that, you don't have to have a randomized controlled trial to meet the moderate quality of body of evidence. You would not give it a High, but you would give it Moderate and, if

there is sufficient number of observational trials -- let's say five-plus -- then the quantity would be voted as High.

So you have options that have been given to you. Did you put yours down? You changed your mind. So now Dr. Rivers. Helen is next. I'm sorry.

DR. BURSTIN: Just a brief process point. So we recognize you need to evaluate, just like everything else we talked about. You have to evaluate what you have in front of you today. We recognize there is always emerging evidence in many of the measures that we look at, but I think you need to look at the evidence as it stands today.

We do, however, have an ad hoc review process. At any point in time, if there is either a material change to the measure or the evidence changes, we will immediately re-review a measure.

So I think, just a process that we have to look at the evidence as it stands now.

CHAIR SEPTIMUS: Thank you. Dr. Rivers.

DR. RIVER: Just a couple of points. When we did the original trial in 2001, we could not ethically have a control group. So we had to put central lines in that group, because that was considered the standard of care. So we never have seen the bottom of what you do -- what would happen if you allow a patient to have what we call standard care. So that trial did not address what we call wild type or standard care.

The other thing is that when you look at these clinical trials, over time -- and this has been since 2001. A study published in CHEST just last month looked at the mortality from 2007 to 2012 in severe sepsis and septic shock nationally, looking at the Medicare/Medicaid database, and the mortality has gone down 12 percent.

So if you conduct a clinical trial over time, whether it is randomized or not, there

is inherent changes in the baseline mortality that may take away your treatment effect. So if you look at a drug like a recombinant activated protein-C, was done in 2001, randomized prospective trial showed that digress administration decreased mortality by six percent.

That trial was reproduced just this year, reported in January this year. The trial is a negative trial, simply because it was done in a lower risk patient population, and it was technically invalidated. So the drug was taken off the market in two randomized prospective trials.

So when we look at trials like that, you have to understand, over time you diminish a treatment effect, and just at the conclusion of a trial -- which 2008 is when these trials started. So we are looking at six years of trial conduction, and you are seeing mortality drop. What does it mean at the end of a trial like that?

So I think it is important to understand that these are not necessarily the end-all questions of the answer to the issue.

CHAIR SEPTIMUS: This reminds me of the Voltaire comment that perfect may be the enemy of good here, but David?

MEMBER SPACH: One question that gets back to the idea of who may be opposing this and how consistent this is with other national recommendations. One of the things I haven't heard is what are sort of the national panels and sepsis guideline panels recommending, and how consistent is that with the bundle that is being proposed here?

MEMBER OSBORN: That is a good question. So pretty much universally the bundles that are being recommended in guidelines mirror the bundle that is being presented here.

To answer the other question that was asked a second ago regarding the randomized controlled trials, every one of them has a control group, and they have a treatment group.

The control group is without the treatment, because they felt in those particular countries there was enough -- it wasn't what people were currently doing.

CHAIR SEPTIMUS: Aaron, and I think we are going to vote.

MEMBER MILSTONE: Sorry. You said "pretty much." Could you just clarify the difference?

MEMBER OSBORN: I'm sorry. I don't remember.

MEMBER MILSTONE: You said the bundles are pretty much the same.

MEMBER OSBORN: I'm sorry. I should be specific. Okay. I'm sorry. It is a very valid point. Thank you for bringing it up.

The international and national guidelines that I have seen are exactly the same as the bundle that is being presented, minus potentially the second draw of lactate in some versions, and the people here who represent the

Surviving Sepsis Campaign -- Although I said I have been involved in that process, they would be the ones to speak to that more clearly. But the newest version of that has to do with one different piece that was not in the Rivers study, which was normalization of lactate.

So, yes, in essence except for the normalization of lactate, the current Surviving Sepsis guidelines mirror exactly what he has put forward.

CHAIR SEPTIMUS: Okay. I think we are going to go ahead and vote. I think everyone, I think, has had a say, and we are probably not going to change too many minds. So it sounds like the undecided. So let's vote on evidence.

MS. KAHN: Voting on evidence, 1c, 1 for Yes, the body of evidence meets the guidance for quantity, quality, and consistency; 2, no, the evidence does not meet the guidance for quality, quantity, consistency; or 3, no, there is insufficient

evidence -- or insufficient information submitted to rate the quantity, quality, and consistency.

CHAIR SEPTIMUS: Remember, this is a stop vote. So if you vote no, then we don't go to the rest of the elements. Okay? So let's vote.

MS. KAHN: We have 11 Yes, the body of evidence meets the guidance; 5 for No, the evidence does not meet the guidance; and 4 for No, there is insufficient information submitted.

CHAIR SEPTIMUS: It is close, but it does pass. So we will go on to the next one, which is opportunity. Hopefully, things will go faster now.

MEMBER OSBORN: So just to make sure that I am saying this appropriately, when we are talking about opportunity, we are talking about the performance gap. Correct?

DR. WINKLER: Yes, we are talking about the performance gap, the opportunity to

drive improvement through use of the measure and other quality improvement activities.

MEMBER OSBORN: Okay. So regarding the gap, as stated before, there are an estimated over 750,000 cases of severe sepsis a year. There are 400,000 that require ICU admission, and there is a significant cost to that.

Looking at the Surviving Sepsis Campaign data, which there are other people who would probably be better to speak to that than me, but all components of the bundle were implemented or were completed in around a quarter or 25 percent of the time.

So that would provide a significant opportunity for improvement.

CHAIR BROTMAN: Any discussion on that matter? All right. Let's go to a vote on performance gap.

MS. KAHN: Voting on performance gap, it is 1 High, 2 Moderate, 3 Low, and 4 Insufficient Evidence. You can go ahead and

start.

Seven High; 23 Moderate; one Low; and zero Insufficient Evidence.

CHAIR SEPTIMUS: Okay. I think we are going to go to reliability and validity. Tiffany, do you want to make any comments about that?

MEMBER OSBORN: I am just looking over our criteria here. So back with reliability and validity, I think that we have discussed the various components of this already. I am open to any questions, if anyone has any further questions.

CHAIR SEPTIMUS: So I think we vote on these separately. So we will start with reliability, which is another must-pass, and the criteria is shown on the screen. Peter?

MEMBER HAVENS: Is now when I can bring up my question about the denominator? Thank you very much. Yes, now it is? Oh, thank goodness.

So the bundle demands placement of

a central venous line. The denominator excludes people who don't get a central venous line for a variety of different reasons. Many of those reasons are coincident with the severity of their sepsis.

This sets up a way in which it makes it difficult to understand how you can apply this perhaps most broadly. Dr. Rivers, could you help me understand the exclusion in the denominator for patients who can't get a central line?

DR. RIVERS: Yes. First of all, to be clear, that is a process of check your lactate, antibiotics, blood cultures, fluids, and once you reach what would be called septic shock criteria, then the central line is entertained.

So everybody who is eligible for the bundle won't necessarily get a central line, because they will basically get better before that. When you get to that point and they require a central line, as a clinician you do

risk assessment.

You say is putting this line in a clinical benefit to the patient over the risk, and that is done with any procedure. So if a patient has a coagulopathy, yes, you do not put the line. That is clinically done for any patient.

So if a patient comes with a heart attack with a cardiac catheterization and you can't get into the artery, that patient has not got a cath.

So that is the reason why you won't have 100 percent compliance, and when you create a quality program, you make accounts for those patients and, therefore, don't penalize a clinician for not doing that procedure. That is what we call standard, the reality of clinical practice.

MEMBER HAVENS: What concerns me is the way that this is written may not penalize the practitioner, but it may penalize patients, number one.

Number two, the ability to put in a central line may be almost as much a marker of the level of activity available in the emergency room or the hospital rather than a specific marker of a medically indicated procedure, and the way this is written will never allow you to look at that.

For example, as I understand your initial study, if the hematocrit was less and the central venous 02 sat was lot, then a blood transfusion would be given in that context.

Is it possible that remeasuring the lactate in the absence of a central venous line and following criteria that might be based on persistence of a low lactate and presence of a low hematocrit would give you equal benefit in the absence of the central venous line?

If that is possibly true, then how can we exclude people who can't get a central venous line, if what we are really trying to do is save the lives of people who have sepsis and septic shock?

DR. RIVERS: Very eloquently stated, but if I can add a couple of facts. Number one, there is no standard. There is nothing. So when you go into somebody's hospital and you say, well, how do you treat a septic patient, there is no standardization up until 2001.

Secondly, patients don't make lactate that can be in septic shock. So a lactemia in septic shock is very common. So if you want to have a patient who comes in with 40 mcg of Levophed and hypotensive, they can have a normal lactate. So that makes lactate clearance not appropriate for that patient.

So I believe lactate clearance is good. So if you see a lactate clearing, that is a good sign that you are going in the right direction, but it is not a standalone methodology for uniformly resuscitating all your patients.

So that is why it is a combination of variables, and I would like to say that you

don't need a central line in these patients. That would be fine, but the key point is you have to take in context that these are a very heterogeneous group of people who come in with a whole number of issues, and what you have to do is attack that patient early and very aggressively with expert -- and expertly, to prevent the downstream effects such as mortality and morbidity.

I must emphasize that this is not a emergency department measure. This is a hospital measure, and the reason why we don't have the adoption of sepsis as we do with heart attack, strokes, is because the professional societies have gotten around those diseases and advocated and made sure that they become hospital system approaches to those disease management. That is what we are asking for with sepsis.

MEMBER HAVENS: I agree wholeheartedly with everything you have said, absolutely. My specific question concerns the

exclusion of people in the denominator for what can either be a marker of a hospital or health system practice approach or an illness severity reason, and it obscures the interpretation of the measurement of the population that we are studying.

This is Section 2 where we are trying to understand the reliability and the validity of what we are measuring. This obscures our ability to understand who we are really measuring, unless this is --So I am frozen here, because I don't know whether -- because there is a huge amount of discussion.

I'm old. When I trained, if you didn't have a Swan-Ganz catheter, then your doctor was a dope, and as a medical student I put in a lot of Swan-Ganz catheters that 20 years later everybody said I was a dope for having done it.

So we have to be careful when we say that this is demanded, and especially if we are going to take people who can't get measured out

of the measurement. Does that make sense?

We are making it so we can't follow it, if we exclude them from the denominator.

DR. RIVERS: I perfectly understand that, and just let me emphasize, the PA catheter is an excellent tool. It is used -- Basically, to use it, that's the problem, and it is not you. But first of all, this is not -- This is standard practice here.

So if a patient comes in or has a line or can get it --

MEMBER HAVENS: But a Swan-Ganz was standard practice as well. All I am asking for is to make it so that the denominator -- we could include people who did not get central venous monitoring in the denominator and, if we pass this now, for you to consider that in the future as a way to identify biology independent of physician practice or health care availability.

DR. RIVERS: Oh, I perfectly agree.

MEMBER HAVENS: Once you include a physician practice capability, putting in a

central venous line as a marker of a population to study, you dramatically change the disease that you are actually looking at.

So what you are looking at here is people who went to a hospital with sepsis or septic shock and there was a practitioner there who could do it.

If that is what you want to study, that is what you are studying. If what you want to study is septic shock, then that is not what you are studying, and so you need to think about changing the denominator, I think.

DR. RIVERS: Very well said. The only thing I can say is what we are trying to do is push medical practice so that every hospital will have that expert in their hospital. Whether it is the emergency department or ICU, that expert will be there to accommodate this patient.

CHAIR SEPTIMUS: Let me -- In fact, if you make it a measure, it is amazing how many institutions will then find people that are

capable of inserting these lines in a timely manner. I will just tell you that, from personal experience.

MEMBER HAVENS: Yes, sir, absolutely. I couldn't agree with you more, but until this specific part of the bundle is proven to be important physiologically rather than for the kinds of -- what it may represent about the health system, it makes me concerned that we don't put it in.

CHAIR SEPTIMUS: Unfortunately, we have already voted on the science part. Let's go. Aaron and then Tom, and then Tiffany.

MEMBER MILSTONE: I had a question on validity, and this, hopefully, is a simple question for you, just a clarification of the denominator.

I was reading in the measure the difference between severe sepsis and septic shock, because clearly, if you had severe sepsis, you only had to get the first four criteria. If you had septic shock, you have

to get all seven, and many of them kind of meet the lower down criteria.

So I wonder if you could just kind of make sure I understand. How easily can people capture that distinction. Right? Because you don't want someone who only has severe sepsis to get dinged for not having lactate remeasured or not having CDP measured.

It looks like there is just some circular stuff about how you define this tissue hyperperfusion. It says: Severe sepsis is defined as systemic manifestations of infection plus sepsis induced organ dysfunction or tissue hyperperfusion.

Then later on under septic shock, you say septic shock is defined as sepsis induced hypotension that persists, and you say sepsis induced tissue hyperperfusion is defined as either septic shock -- So there is a circular -- What you talk about is tissue hyperperfusion.

So I just want to make sure you can include how you clearly distinguish patients

with severe sepsis versus those with septic shock. So that is a definitional issue.

The second thing is how well can practitioners or people do that in terms of a measurement, like the validity of people being able to distinguish those two populations.

So I think no one would argue can you identify patients with sepsis, but can most hospitals distinguish after the fact in terms of compliance measurement those two groups, sepsis versus septic shock?

DR. RIVERS: Very good question. The key point is hypotension refractory to fluid administration basically says that you require a vasopressor. So you are in septic shock, period.

So the idea of having a hypotensive patient on pressors requires that you have advanced monitoring. So that is a risk stratification. It puts you into a mortality of 49 to 50 percent.

If you have persistent lactate

elevation -- so if your lactate comes in and it comes in at greater than four after you have done your fluid challenge and after you have given the patient antibiotics and initial resuscitation, you equivalently have the same mortality.

So you are dealing with a high mortality with a lactate greater than four or hypotension after fluid administration.

MEMBER MILSTONE: So when all is said and done, maybe someone wants to look at the 2a1.7 and see if that -- So I agree with you clinically. I am just trying to decide whether that is how I would interpret this from distinguishing the populations.

DR. RIVERS: I'm sorry?

MEMBER MILSTONE: You just said that patients who received vasopressor support, by definition, have septic shock, but it doesn't mention the definition of the use of vasopressor support.

DR. RIVERS: Clinically, if you are

hypotensive and not responding to fluids, then that --

MEMBER MILSTONE: I agree with you, but I am trying to get at the definition that is listed, and it is 2a1.7 on -- I am not sure what page it is -- on page 20.

DR. RIVERS: So your question -- I am just trying to --

MEMBER MILSTONE: I am just trying to decide: When you are going through, do you need to meet the first four elements of the bundle or the first -- or all seven elements of the bundle? How will those patients be distinguished.

You say clearly in the front that it has to do with whether you have severe sepsis or whether you have septic shock. I just want to make sure I understand clearly how clinically those are distinguished, based on these criteria.

DR. RIVERS: So if you have suspected infection and you come in hypotensive

and you respond to fluid, then you basically are now a severe sepsis patient. So you don't have to go around to push the bundle to completion in terms of central line placement.

The people who persist, that require a central line, are patients who have persistent hypotension. So if I give you four to five -- three to four liters of fluid for an average seven kilogram person, and you are hypotensive, pressors are not written in there, but that is the clinical reaction, is to use pressors.

So I am not trying to -- You know, lactate greater than four, uniformly, if you look at articles by Steve Trzeciak, Nate Shapiro, you look articles out of University of Pennsylvania, 3,000 studies show that if your lactate is greater than four, your hospital mortality is anywhere from 28 to 50 percent.

MEMBER MILSTONE: I agree with you. I am just saying, does it somewhere in here specify what the criteria are for septic shock?

Just the average person, how will --

DR. RIVERS: I understand that, and if this -- the typo and the way it is written, we can -- But it is basically the same thing, the Surviving Sepsis Campaign recommendations. So however this is transposed, I can -- Sure.

MEMBER MILSTONE: I think that would be really an important thing. I mean, that is defining the population that have to get 5, 6 and 7. I think there would need to be clear criteria for validity. Otherwise, my hospital might interpret septic shock differently. They might say, oh, well, lactate is not important or -- I am not saying they don't, but I don't know if it is that clear to whoever is going to be assessing.

CHAIR SEPTIMUS: Maybe I am not reading the same document, Aaron, but it says clearly here, in the event of persistent arterial hypotension despite fluid resuscitation (septic shock) or initial lactate of greater than or equal to four millimoles,

measure CVP and measure SCV-2.

MEMBER OSBORN: Aaron, are you asking how it is being measured, reported?

MEMBER MILSTONE: I just think somewhere it should be stated that there is a clear definition of patients that meet criteria for septic shock. You have provided them. I am just not finding it.

Yes, I understand like in the introduction, it does say --

MEMBER HAVENS: On page 1 of the initial document, 2a1.1, the numerator statement of the initial 005 big document that was sent out, is the mean arterial pressure that identifies hypotension initially not responsive to fluids.

MEMBER OSBORN: I think it is almost like we need to be sure that the definitions of septic shock and sepsis are clarified a bit, because -- Alexis, scroll down to where you were before, because here are the measure specs, and here the denominator details. That is where

we put the definitions and all the things that need to be crisp and clear for all end users.

So I think, Aaron, is this where your question is?

MEMBER MILSTONE: Yes, I guess these are -- Thank you for pointing out the one above. I think those do give guidelines as to how you are -- These give responses to these values. So this is saying, if your patient has an initial lactate of better than four, you should do this; if your patient-- So I think you are outlining it here. I just don't think it translates down to the denominator where, if you included these -- So is your denominator including patients that are on vaspressors, patients who have initial lactate of greater than four.

I think, if that is your denominator, that would be very easy for people to interpret, moving forward. I just didn't see that as in the denominator for very clear criteria.

DR. RIVERS: Yes, I understand.

CHAIR SEPTIMUS: But if we made sure that was clear, that would meet your concern?

MEMBER MILSTONE: Yes. I hadn't -- Yes.

CHAIR SEPTIMUS: Okay. Tom?

MEMBER FILE: Actually, I had several comments here, and I am not sure many of these go to feasibility, but I am going to make some of them now, and they relate to --

CHAIR SEPTIMUS: We are not talking about feasibility. We are talking about reliability.

MEMBER FILE: I know, I know. Well, yeah, but I am looking at the criteria here that -- where in heck was it? -- that all information required to identify and calculate the target population denominator, such as definitions, codes with the descriptors, and/or specific data collection items and the responses.

I think my comment is somewhat

similar to Aaron's here, is that -- and let me just ask our NQF collaborators here or colleagues. When Emanuel gives these criteria, I also had questions about specifying exactly how patients would fit into these criteria, and is this going to be variably interpreted by data collectors, because you say fluid resuscitation. Well, how much fluid.

You said four liters. I mean for a 70 kilogram person. Is there a specific criteria of how much liters? The denominator here is a clinical criteria, set of criteria. Now it is somewhat different than when we were talking about the prior two measures. In fact, at this rate, we are going to be -- whatever. But at any rate, where we were talking about specifying an ICD-9 code, and I can understand what Peter said before: How well are those ICD-9 codes -- do they correlate truly to the diagnosis that we are trying to capture?

Well, here you are talking about a clinical constellation of manifestations, and

I just don't know -- ICD-9 codes for sepsis and sepsis with SERs and shock. I mean, you could do that, and I don't know how well that correlates with this.

I can see how this could be very valid in a study setting where you have got investigators looking at all of this data within the charts. I just don't know how -- and it goes back to what Ann was saying. Just precisely defining this population so that data extractors, who probably aren't going to be as expert in this field, obviously, as you guys -- and by the way, let me just say I admire all the work you and Tiffany have done on this, and we appreciate it, and I hope that when I get septic shock that you guys will take care of me. But nevertheless --

So that is one of my concerns, is how well do you think that this can be valid in a measure for a data extractor who is not expertise?

I have some other comments. I think

I will delay those. They are feasibility. But then the only comment I had right now about data extraction and defining specific data elements is you say the administration of broad spectrum antibiotics.

Now I talked about this in our Work Group, and my only comment was I don't know what broad spectrum antibiotics means. So like for example, for our pneumonia measure, we actually gave a list of antibiotics that would be appropriate for severe pneumonia, which would be, quite honestly, one of the more common causes of sepsis, and that is severe pneumonia requiring ICU admission.

So we actually give what would be appropriate empirical antibiotics and, if you don't use any of those antibiotics, then you are in variation of the measure.

So I don't know what broad spectrum is. Moxacin sounds broad spectrum for some people. And again, it is just relating to who -- for definitions of obtaining this information

is my comment for validity.

CHAIR SEPTIMUS: Tom? I'm sorry, Tiffany, then Tom.

MEMBER OSBORN: This actually is related. I originally thought this was going to go into feasibility, which is why I didn't bring it up, but since the conversation is going this way now, there is a contingent that is concerned regarding how the components are defined.

So this is a timed measure, and it is not clear how time zero is defined. So if you have a 35-year-old who presents with uncomplicated pneumonia to the emergency department that developed shock three hours later, and maybe that patient is still in the ED, maybe they are in the hospital, how is time zero defined? Is it triage time, and people want -- This contingent doesn't want to be held accountable for addressing something that didn't exist at the time they saw the patient. That is one thing.

When you talk about the ICD-9 codes, the question becomes, well, is it the ED ICD-9 code? Is it the hospital ICD-9 code or the discharge ICD-9 code, and how is that impacted by where the patient develops severe sepsis or septic shock?

I know that there are health plans and integrated delivery systems that are currently implementing versions of this bundle in unique ways, and I would be interested in hearing about those, but the question would also come, whether or not those unique methods will translate effectively to urban, rural, academic or community settings.

Some would advocate, as has been spoken a second ago, that these detailed implementation specifications should be brought forward and discussed in another steering committee or available in a form for review and public comment by stakeholders. So that question has been put to me via email. So I wanted to make sure that that was brought up.

CHAIR SEPTIMUS: Tom?

MEMBER GIORDANO: Yes. Looking at what has been submitted for reliability and validity, I am not convinced that there is adequate evidence or high quality evidence to support that these measures as written are reliable or valid.

I have a question for the NQF, which is: Is it expected that this document or a modified version of this document would be all you would need to operationalize these measures? In other words, I see a lot of ambiguous and vague statements in the numerators and the denominator statements. Is it expected that you could operationalize based on this document?

DR. BURSTIN: This is Helen. Just to speak to the first issue, I think if you look at 2a21, there is actually -- This is actually quite a bit of testing, 498 charts reviewed by nine independent abstractors. That is actually quite high level in terms of evidence of reliability, and there is significant evidence

in terms of validity.

It sounds like -- We were just conferring. It sounds like there is an entire attachment that I am not sure went through to you that has the very detailed specifics, and perhaps we will make sure that gets to you.

MS. BOSSLEY: Right. We will get it to you, but I am looking at the data collection tool. It is a sample data collection tool that walks through. It appears, I think, as if it is someone doing a paper medical record extraction step by step on exactly what you would look for that matches the specifications that Reva showed you.

So I think we need to get this to you, because it sounds like it is going to help answer that question of how you go about abstracting that data, and then we have got the reliability testing data that fits under 2a2, as Helen mentioned.

MEMBER GIORDANO: Perhaps that should have been supplied earlier. I mean, this

process is supposed to be a fairly rigorous and objective review of the data, and I will go on the record as saying that, if there are important elements how to operationalize these that are not adequately defined in the document that we have, then I think to not provide us with some evidence that these can be operationalized and have been operationalized successfully is maybe not fair to us. We might have been able to cut this discussion in half, if not more.

I also don't know that a study in one health care system adequately -- which is what the element that you pointed out -- adequately addresses reliability and validity. It may be that Henry Ford Hospital system has it down perfect. I mean, they developed it, and they are doing a great job with it, but does that mean it can translate to other health care systems?

CHAIR SEPTIMUS: Emanuel?

DR. RIVERS: We provided what we call usefulness for public reporting of the

measure. There is Kaiser Health Care System, which probably has done over 8,000 patients. San Francisco Hospital Coalition took 11 hospitals over three years, conducted the same process improvement. Catholic Healthcare, West Center Healthcare, again Noma Linda University, University of Kansas over 7,000 patients in the last three to four years, same collection tools, etcetera.

MEMBER GIORDANO: Are those in here? Is that summarized?

DR. RIVERS: Yes, on page -- It is in the summary under usefulness of public reporting. We provided all institutions that have -- Intermountain Healthcare in Utah, same outcomes. So multiple institutions, large scale institutions have done the same thing .

MS. BOSSLEY: At page 29 of your PDR.

MEMBER GIORDANO: I am sorry to perseverate on this, but that is usefulness. That is not saying that they get the same --

If other people look at it, they come up with the same results. That is not reliability and validity. So maybe it is extra analyses that need to be done to say that those are also reliable and valid, the measures in those various systems.

MEMBER BRADY: I just wanted to add to that, that it says in the description as well that the reviews were done by nine different clinicians, and I think in terms of -- maybe this is really a feasibility and usability issue -- that that is a high level reviewer. I don't think that that necessarily, when this gets put into practice, who is going to necessarily be doing medical chart reviews.

DR. RIVERS: Well, in reference to that statement, that was to basically validate. We have a sepsis coordinator who examines all of our septic patients, and to test her and make sure that she has sound validity in our patients, we do what they call a back-physician analysis, and that was that representation.

So once every year or so, once every two years, we get a group of charts. We all go through them as physicians, because we -- and basically validate that she is doing the correct thing, and that is what that was. It was a validation, which is basically a quality check.

CHAIR SEPTIMUS: Let's take one or two more comments and then go for a vote at this point. Tiffany?

MEMBER OSBORN: Again, I am trying to be fair to the multiple different comments that I have received prior to this. So in relation to that, some would advocate that these detailed specifications should be brought forward in a steering committee so that the appropriate stakeholders could comment in a meaningful way, and that has not been done at this point in time.

CHAIR BROTMAN: Tom, did you have your --

MEMBER OSBORN: Okay. The people

who would comment on the importance of these specifications would say, yes, there may be certain health systems or integrated delivery systems that have looked at this independently or individual institutions that have looked at this individually, that there are more stakeholders that are involved in the process than that, and they might want to have -- they might have valuable input to put into that discussion, and that the NQF would be the appropriate place to do that.

DR. BURSTIN: I agree completely, but that is why this is a process. You guys are actually the earliest part of the consensus process. Following your recommendations will be a 30-day comment period, and they will all be very welcome to see it in all its glory. We will include the full appendices, and we would welcome comments. I suspect we will have many, as we often do on measures that are a bit controversial, but that is what the process is intended to do. You are the first step on this

process.

MEMBER OSBORN: So just to be clear then, if you had stakeholders who said that there were ways in which they felt the way in which these specifications were being done needed to be revised, then how would that go forward?

DR. WINKLER: These specifications are not being presented to us.

CHAIR SEPTIMUS: I am going to piggyback on what Helen said. We go through these measures based on what has already been presented. We vote. We have a time for public comment several times during this meeting. If the measure is approved, it will get posted for public comment. Then we will see those public comments and make any revisions or changes that need to be made.

So our goal here as a committee is really to look at the information that has been presented to us, not the -- The other stakeholders are going to get a chance to comment on this. So let's not mix the groups up.

Manny, you had one other thing you wanted to say, and then we are going to go vote?

Okay. So let's vote on reliability, and this is a stop measure. So if you will read the measure.

MS. KAHN: Yes. Voting on 2a reliability. It is 1 High, 2 Moderate, 3 Low, and 4 Insufficient Evidence. You can start now.

We have 1 High, 7 Moderate, 5 Low, and 7 Insufficient Evidence.

CHAIR SEPTIMUS: Okay. This measure failed. So we stop here. It fails. So we stop here, and we don't go on to the other parts of this measure.

We, obviously, are running slightly behind, and what we thought we would do is we would now ask for public comment, and then after public comment, we will break for lunch. Then we will get to the hepatitis measures after lunch.

So, operator, we are going to take public comments.

OPERATOR: At this time, I would like to remind everyone, in order to ask a question, press Star, then the number 1 on your telephone keypad.

CHAIR SEPTIMUS: And of course, anyone here in the room who would like to make a public comment as well.

OPERATOR: Your first question comes from Jeremiah Schuler with ACEP.

MR. SCHULER: I am the incoming Chair of the Quality and Performance Committee for ACEP, and I will keep this brief, because this regards the sepsis measure which just did not pass.

On behalf of ACEP, we had concerns about the last category around reliability and validity, and we hope that we can work with the other societies in the Surviving Sepsis Campaign to fully specify this so that there is data on reliability and validity, because we feel that this is an important topic for which there is good quality improvement evidence, and it would

be appropriate for there to be a measure, but that it needs to be fully specified and then tested. Thank you.

CHAIR SEPTIMUS: Thank you.

OPERATOR: Again, to ask a question, press Star, then the number 1 on your telephone keypad.

At this time, there are no further questions.

CHAIR SEPTIMUS: If there are no further questions, we will -- Are there any --

MEMBER OSBORN: Yes. The comment that I had is that --

CHAIR SEPTIMUS: This is public comment.

MEMBER OSBORN: So we don't -- Okay.

CHAIR SEPTIMUS: It is public comment. So we are going to break for lunch. We were scheduled to come back at 1:15. We are already running behind. So I think we ought to -- I think we are going to have a slightly working lunch.

We are going to come back at 1:15 and start.

(Whereupon, the above-entitled matter went off the record at 12:48 p.m. and resumed at 1:16 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

(1:16 p.m.)

CHAIR SEPTIMUS: The next three measures are the PCPI. Dr. Wong is here to help guide us through these three measures. That is 0399, 0400, and 0393 are from the AMA. So, John?

DR. WONG: Thank you, Ed. I'm John Wong. I'm a general internist at Tufts Medical Center. I am Chief of the Division of Clinical Decision Making. I am one of the co-chairs for the Hepatitis C Workgroup. And it is my pleasure to be here on behalf of my co-chair, John Ward, who is Director of the Division of Viral Hepatitis at the CDC.

Also here but out in the hallway is Mark Ghany, who is a member of the workgroup and also at the NIH. And I am joined by staff members from PCPI.

So just to give you some historical background. Around 2004, the American Association for the Study of Liver Disease, the

American Gastroenterological Association and the AMA through the Physician Consortium for Practice Improvement or PCPI formed the Hepatitis C Workgroup. The initial quality measures were approved by the PCPI in 2006, updated in 2008 and reviewed and updated again just this past June.

Nine of nine measures were recommended for full endorsement by the NQF Consensus Standards Approval Committee in November of 2011 and they are currently being reviewed for endorsement maintenance with your group.

I will just point out that all nine of the submitted measures have been tested for reliability and validity and are currently in use in CMS's PQRS program and I will say a little bit more about that.

I wanted to speak briefly about several of the measures. In particular, since bundling came up in the last extensive discussion, I want to explain why the hepatitis

A and hepatitis B vaccination measures are bundled. It seems self-evident if you are trying to protect for hepatitis A, you should also protect for hepatitis B.

The second bundled measures are the measures for checking or confirming that the patient is still viremic prior to treatment and secondarily identifying the genotype. Those are both important, obviously, because if the patient is non-viremic they don't require treatment. And genotype is very important for determining the particular kind of treatment and the duration of treatment.

I want to focus more extensive comments on things that we spent a fair bit of time talking about in our PCPI workgroup and that is measure 0397, having to do with treatment at a minimum with pegylated interferon and ribavirin. And then afterwards, I am going to turn to 0398, where the language is no greater than -- checking a viral load at no greater than or equal to 12 weeks.

With regard to establishing a minimum treatment, the workgroup decided to try to balance the measure burden, along with the absence of typically having test results. So we well recognize that we are on the cusp of an explosion of new hepatitis C drugs with over two dozen drugs under development currently and multiple different kinds of regimens and we anticipate substantial changes over the next two to five years.

We, however, elected to stay with a minimum of peg plus riba because it would not require the specification or the quality measurement to know exactly what type of genotype that particular patient had. Although some EMRs have that available, some system levels have that available, the majority of them don't have that.

In addition, there are some clinicians who even though the current standard of practice is triple therapy for genotype 1, some clinicians are treating with just pegylated

interferon and ribavirin after they do IL-28b testing because in the group that is CC positive, pegylated interferon plus ribavirin has been shown to have the equivalent sustained viral response rates and by testing, they can avoid the side effects of the protease inhibitors. In addition, they incidentally happened to reduce the cost of therapy by two-thirds.

I want to turn now to Measure 0398 where again we had extensive discussion about how and when to measure viral response to antiviral therapy. And we again elected to establish what I would call a low bar measurement. That is that somebody assesses the viral response at 12 weeks or before that. We recognize that extended viral response, rapid virologic response are all part of the initial criteria for optimal treatment. But we wanted to decrease again the measurement burden on users to demonstrate quality improvement or accountability.

And we also noted that as a single

measure regardless of genotype, it covers all of those and in addition would cover the two new protease inhibitors where again testing differs depending on which protease inhibitor you are using and, consequently, the language for the measure would have to be specific for the treatment and for the specific time at which you measured response. And again, I think the phrase was perfection is enemy of a good --

If I could just spend a few minutes just talking about the importance of the measures. Current estimates are that 3.7 to 4.1 million Americans have chronic hepatitis C, if you include those who are incarcerated or homeless. It is the principle cause of death from liver disease and is the leading indication for liver transplantation. Projections from the CDC suggest that 1.76 million people will die from -- will develop cirrhosis and that another 400,000 will develop hepatocellular carcinoma. And in the absence of treatment, one million people will die from hepatitis C.

To corroborate that evidence, hepatocellular carcinoma is the fastest growing cause of cancer-related mortality in the United States and hepatitis C accounts for about 50 percent of those cases.

I am going to turn now to the performance gap. And I just wanted to -- I think you have all been provided with some data from PQRI or PQRS. And I want to point out that that represents only 24 percent of eligible professionals. And I will also point out that for the most part, those are professionals who volunteer to report their outcomes. So there is an emphasis on performing those measures and, as such, if they do perform to those measures, they get a boost in their pay and so I would submit that that is probably a slanted perspective of current practice. And, in fact, when multiple publications, in particular the one in the *Annals of Internal Medicine* have looked at that, performance is underperformed.

I also want to highlight the recent

CDC announcement just 11 days ago which has advised the screening of the birth cohort that is Baby Boomers born in 1945 to 1965. As such, I think many physicians will be checking for hepatitis C who previously may not have because of the publicity assigned to that. And as such, they may be less familiar with the quality measures that we are talking about.

I briefly want to mention things about reliability and validity. As I mentioned, these measures have gone reliable -- have been demonstrated to be both reliable and feasible. In fact, they have face validity with a survey and expert panel rating of the validity statement, where they agreed or strongly agreed with these measures. And in particular, the annals article that I mentioned tested these measures in 14 million members in multiple data sets, including extensive detailed clinical data.

I briefly want to mention measure development process. We, as you, have a

cross-specialty, multi-disciplinary workgroup that includes all medical specialties and allied healthcare professionals. In addition, we try to include members of lay organizations, including patients, consumers, private health plans and employers. We rely on clinical practice guidelines as the foundation for the development of performance measures, based on their evidence review. As you all know, the Institute of Medicine has raised the bar for the development of trustworthy guidelines and, as such, we provided you with a summary of the literature to supplement those evidence reviews in the guidelines to help you evaluate the quality consistency and validity of the data.

With regard to usability and feasibility, our measures undergo extensive public comment and peer review processes. And also have very precisely defined technical specifications, keying in on electronic health records and, in particular, Category 2 CPT codes, which will facilitate administrative

coding of the quality measures.

In summary, our workgroups sought to focus on those areas with the most potential for impact, where there was the strongest consensus about the best practice, where the likelihood for unintended harm was lowest. Moreover, the group sought, as much as possible, to keep the measure straightforward, trading off the measurement burden and the quality improvement opportunities aligned when appropriate with measures developed by others and clinically sensible, giving the clinician the latitude for judgment about the appropriateness of an intervention when such latitude is justified. Thanks.

CHAIR SEPTIMUS: Thank you, John. That was excellent.

Okay, we have 0399 and 0400, hepatitis A and hepatitis B vaccine. Curtis, I know you have the A and then we will go through that measure and then Mohamad will go through the hepatitis B vaccine but they are very similar

and paired measures. So hopefully the discussion will be hopefully the same.

MEMBER COLLINS: Yes, thank you Dr. Wong, for the wonderful introduction. I think he covered a lot of what we are all going to say here and we are thankful for having you here as a resource.

So measure 0399 is the percentage of patients 18 years and older with a diagnosis of hep C who have received at least one injection of hep A vaccine or who have documented immunity to hep A.

And I guess as we go forward, the impact or justification for this, the developers listed a lot of statistics which we just heard for hepatitis C. However when we asked about this in the workgroup, they got back to us and said there are really no specific data available on the incidence of hep A and patients with chronic hep C. But the argument is that vaccination decreases potential for patients acquiring hep A, which could contribute to

further liver damage. So I think that kind of sums up potentially the impact.

CHAIR SEPTIMUS: I think also, on the call, it was discussed that people who get acute hep A or chronic hep C are the ones who have the most significant morbidity and potential mortality as well.

CHAIR BROTMAN: Any discussion on that point?

MEMBER HAVENS: So are the data presented here for that or not?

MEMBER COLLINS: So I guess I was not aware of any data presented here as far as that goes, no.

MEMBER CHUNG: There is some supplemental data that you guys presented. Right?

MS. WINKLER: Well, I think -- let's go in order. If we are talking about impact, okay, then the information presented under 1a.3 and estimated 180 million people are infected worldwide. Between 1999 and 2002 1.6 percent

equaling four million persons positive for hepatitis C, 80 of whom are estimated to viremic. It is the principle cause of death from liver disease. So that data is in your submission form around impact.

MEMBER HAVENS: Yes, ma'am, but this guideline is about the impact of hepatitis A vaccination in that very large group of people with hepatitis C. We all agree there is a lot of people with hepatitis C. The question is, how many of them are at risk of getting hepatitis A and there are no data on the rate of hepatitis A co-infection presented in this document, number one. And then, while we all believe that if you get hepatitis C and you get hepatitis A, it is bad for you, there are again no data addressing that issue presented in this document.

MEMBER CHUNG: There are supplemental data here that PCPI I think provided us. Correct? I mean, it's the Word file that you guys sent us, I don't know if you

got that, about I think speaking to Ed's point about the mortality of acute hepatitis A superimposed on chronic C. And yes, it is not a precise epidemiologic characterization of the risk of hep A in C but it is the risk of morbidity and mortality in those incident cases of A superimposed on C with essentially seven of 17 patients developing fulminant hepatic failure, six of whom succumbed. That is what Ed was referring to, I believe.

And so I mean I think on a case severity basis, if not numerical, I think an important point about the utility of vaccination in this group of patients could be made to prevent the morbidity and mortality of A, should it occur.

DR. WONG: If I may add, in data that I didn't present but there are other data suggesting that about roughly, depending on the study, you are looking at about half the patients with hepatitis C do not have antibodies to hepatitis A.

MEMBER HAVENS: And then to that point, if they are vaccinated with hepatitis -- in a patient with hepatitis C, what is the efficacy of vaccination for one dose which we are asked to comment on here versus two, the 80 percent coverage after a single dose is not in a population with hepatitis C, as I understand those data. So what are the data that we are asked to comment on actually makes any difference in terms of immunity in this specific patient population?

DR. WONG: So Emmet Keeffe published a paper in *Hepatology* in 1998. It is in that same supplemental email, Word document, that demonstrated both the safety and immunogenicity of hepatitis A vaccine in hepatitis C patients.

MEMBER HAVENS: For a full -- for a two-dose. And after one dose?

DR. WONG: I don't recall exactly but it was close to 80 percent.

MEMBER HAVENS: Thank you.

CHAIR BROTMAN: Doug, did you have a question?

MEMBER CAMPOS-OUTCALT: Yes. Yes, a comment and then a question.

First of all, before I make my comment, because it is good to be interpreted as being negative toward the indicator and I am not. But I can't let data go by like was just presented, which was out of 17 patients, seven died. That is very selective data. I mean, hepatitis A for adults is asymptomatic most of the time. So these are people who were symptomatic and had some sequelae and were discovered. You know, that is true. That is like saying West Nile virus as a fatality rate of 80 percent but it really doesn't because most of the disease is subclinical. So that is not very convincing data.

But having said that, my question is many people with hepatitis C are at high risk for hepatitis A and B and ought to be vaccinated just based on their risk factors. And so is

there another indicator somewhere regarding vaccination that we are not aware of that would already cover this, you know, another quality indicator somewhere were adults at-risk ought to be vaccinated against hepatitis A and hepatitis B?

MS. WINKLER: Yes, actually NQF does have one other measure. Actually they are talking about in the other room in another project. For patients with chronic liver disease, I believe it is the hepatitis A -- I was trying to think if it was A or B -- in all patients with chronic liver disease. So that is the broader population. But other than that, no. That is the only other one.

MEMBER CHUNG: So if both go through, we will be approaching the same problem from two different angles?

CHAIR SEPTIMUS: Actually just one follow-up, Doug. Actually, the rate of symptomatic hep A in adults is much higher than pediatrics. The vast majority of pediatrics

is asymptomatic. In adults, it is close to 50/50.

MEMBER CAMPOS-OUTCALT: So there is a lot of asymptomatic disease. That is my point.

CHAIR SEPTIMUS: There is, but it is much more slow in pediatrics.

MEMBER CAMPOS-OUTCALT: You can't say that the fatality rate is seven out of 17.

MEMBER CHUNG: But the case fatality rate for symptomatic disease is awfully high here. I mean, in this group of patients. Right? Symptomatic hep A does not kill the vast majority of patients who have symptomatic hep A. So this is certainly a high odds ratio.

MEMBER CAMPOS-OUTCALT: Yes, but there is a lot of limitations to it. I mean, I just object to having that data presented as the complete data for mortality rates of hepatitis A and those with hepatitis C. It's not.

CHAIR BROTMAN: Any other comments

at this point? Okay, let's vote on impact then.

MEMBER CAMPOS-OUTCALT: Question.

CHAIR BROTMAN: Yes.

MEMBER CAMPOS-OUTCALT: Are we talking about -- does this impact vote also include the -- when we vote on impact, does it include the already compliance that occurs with an indicator? This one is not a question for this one but for the future ones. For hepatitis C, we have some things already being performed at 90 percent compliance. Is that part of --

MS. WINKLER: Remember, you are going to vote separately on impact, evidence, and then performance gap. And that is more --

MEMBER CAMPOS-OUTCALT: A performance gap?

MS. WINKLER: -- for the performance gap.

MEMBER CAMPOS-OUTCALT: Okay. All right, thank you.

MS. KAHN: Voting on high impact.

High, moderate, low or insufficient. You can go ahead and start.

(Pause.)

MS. KAHN: So we have five high, ten moderate, one low, and four insufficient evidence.

CHAIR BROTMAN: Okay, that passes. Let's go to the evidence.

MEMBER COLLINS: So we've -- it's going, you know, it is going to be tough to keep that from merging over but as far as from the impact to the scientific evidence, as far as the scientific evidence, the developers listed a single report that suggests that superimposed hep A and virus infections in persons with chronic liver disease, particularly those with hep C was associated with fulminant hepatitis. Therefore, it was recommended chronic HCV infections who lack evidence of pre-existing antibodies to hep A be administered the hep A vaccine.

And that is coming from what I

believe is the guidelines. So the level of evidence for the guideline is 2a, level C, with level C consensus opinion of experts level of evidence.

Now this was brought up in the workgroup and we have talked about it already with the severity of disease in patients who do have chronic hep C who had hep A superinfections but there was a comment in there as well and the start of the sentence is, "although subsequent studies have not found comparable morbidity and mortality results." Now we received this late yesterday afternoon. So I haven't had a chance to take a look at these two or three other studies but I was wondering if the developers could comment on those. Because we present the one study that has the high death rate, high rate of complications but then there is mention of two or three other studies with -- subsequent studies have not found comparable morbidity or mortality.

DR. WONG: Correct. It is very

hard to find these studies.

MEMBER COLLINS: Yes.

DR. WONG: These are not systematic reviews or meta-analyses. These are reports typically at the country level or otherwise. These typically are very small studies and as such, may lack some power to detect the same outcomes.

I will say in some of those studies it was not clear -- well, some of the populations in those studies included patients, as you might guess, who were simply antibody-positive. So we actually don't know if they were hepatitis C viremic at the time of their hepatitis A.

The one study that is mentioned is the one that Ray kindly mentioned the seven out of 17 patients who developed fulminant hepatic failure is the one that is the most widely cited and partly because it is in the *New England Journal of Medicine*.

So when people talk about morbidity and mortality associated with hepatitis A,

superinfection, or coinfection on top of hepatitis C, that is the one that everybody points to.

CHAIR BROTMAN: Any other comments? Okay, let's vote on the evidence.

MS. KAHN: Voting on 1c, evidence. You can go ahead and start.

(Pause.)

MS. KAHN: We are missing two votes. Would everyone just press their clicker one more time?

MEMBER BEAL: Yes, you need to speak up. I can't hear you at all. I'm just hearing bits and pieces.

MS. KAHN: Sorry. We are voting on 1c, evidence.

(Pause.)

MS. KAHN: So we have seven for yes, the body of evidence meets the guidance; six for no, the evidence does not meet the guidance; and seven for no there is insufficient information to submit it.

MS. WINKLER: Okay, hold on. I mean, technically the vote is that it does not meet the criteria, which given the discussion, is probably accurate.

The committee at this point has the option to invoke an exception to this criteria, if you feel that despite the lack of the evidence meeting the criteria, it is still important enough, I guess or something, that you want to say this is an exception and it is an exceptional circumstance but we still feel that it should go forward for endorsement.

If you would like to do that, we do have that potential exception to the empiric evidence. Any discussion about that thought? Do you want to go there?

MEMBER FILE: Well my only comment is one who is a big proponent of preventative vaccine use and when you consider benefit, which I acknowledge has not been demonstrated in this particular situation versus harm, which I would consider extremely minimal, I would be in favor

of an exception.

MEMBER CHUNG: I would argue right along with Tom on this one. I mean, this is what we would consider primary care for chronic liver disease patients. This is Harm Reduction 101. We tell our patients about alcohol. We try to reduce future risks of drug-induced liver injury preventing hepatic toxic medications. That is part of our primary care for these patients. The same thing applies to vaccination, even in the absence of iron-clad evidence, substantive evidence for clear-cut benefit over the long haul or over large numbers of patients. I would make the case for carrying it forward.

CHAIR BROTMAN: Thank you. We have got a couple of comments. Let's see if we can go quick. Doug?

Oh, okay, Mohamad, I think you had yours up first.

MEMBER FAKIH: Thank you. I will support Tom and Raymond's comments.

I think the other thing that is very interesting about hep A if you get the vaccine only once, you get 80 percent protection, which is something that we don't see in hepatitis B vaccination and we do this all the time for any travels to endemic areas for a very short period of time. So I think it is very protective. I mean, I think it is something we should do.

CHAIR BROTMAN: Good point. Michael?

MEMBER FARBER: Well the question that I wonder is is if you have the data. I remember the fact that adults over 50 that get hepatitis A have a two percent mortality that went back a long time ago. Maybe that has been challenged. Whereas in children, of course, the mortality is almost zero.

So my question really is is how does the hepatitis C group compare to a group of normals. In other words, should people that don't have hepatitis C, do they have the same risk as hepatitis C to get a fulminant case of

hepatitis A when exposed as adults.

CHAIR BROTMAN: Is that evidence out there?

DR. WONG: So the *New England Journal of Medicine* article was the first to really shine a bright line on this and it suggested substantially increased mortality among those with symptomatic hepatitis A on top of hepatitis C.

In terms of efficacy statement, I will point out that the CDC says it will last 20 years. Hepatitis A cases have declined by over 90 percent. And since the immunogenicity of the vaccine is comparable with those with hepatitis C, I would expect a comparable reduction at the very least in those with hepatitis C from being immunized with hepatitis A.

CHAIR BROTMAN: It appears Dr. Beal has a question. Why are we not voting for or against this recommendation for an exception? That is the next step. So we are there.

So let me just get a couple more comments around the room. Rekha?

MEMBER MURTHY: Thank you. I was just going to comment and add to Tom and Raymond's comments now there is about strongly endorsing an exception to this. I think the other caveat I would add -- not caveat -- but as a point I would add is that unlike behavioral interventions like alcohol counseling, et cetera, this is something that actually can be a specific intervention that doesn't require behavioral modification and even a marginal improvement in risk is worthwhile.

CHAIR SEPTIMUS: Just so we can keep things moving, remember at the beginning of the day, if you have something new to add, we would love to hear from you. But if everything is said that you want to say, then we can move more quickly to votes. So just keep that in mind, as we move forward this afternoon.

Aaron, did you have something you wanted to add?

MEMBER MILSTONE: So as a pediatrician, I am a huge vaccine supporter. I just want to point out and I have never been doing ACIP, you know the Association on Immunization Practices for the CDC, their panels but there is often discussion about cost. And I think in this case you are telling us, on one hand, that there are going to be over a million potentially new baby boomers identified with hepatitis C and here we are coming up with, or at least discussing the idea of avoiding evidence and still recommending universal vaccination of that entire population.

So I love vaccines and I would say every person should get every vaccine out there but it is not cost-effective. So I think I would like to hear a little more. Before I say evidence aside, let's recommend this, I would like to know is there any data on the cost benefit of that. So saying a million vaccines versus seven mortalities. I know it is not that but I think I would need a little more to say there

is a some cost benefit of this versus other measures that are important for other patient populations.

CHAIR SEPTIMUS: Can I ask a question? Helen, in terms of cost implications for the recommendation, can you comment on NQF's position on that?

MS. BURSTIN: I was waiting for somebody to ask that. It is a great question. I think in this day and age it is hard not to look at quality and consider affordability. I think it is reasonable. I don't think it really goes into the evidence question but I think it is, at least, a reasonable thing to discuss but it shouldn't really get factored into the evidence for the measure focus.

DR. WONG: Do you want me to answer the question? Because I happen to know there have been multiple cost-effectiveness studies done in this area. Unless you are from an area with high endemicity where you are likely to be antibody positive just by natural exposure,

multiple studies in the United States suggest that it is cost effective.

I will also mention that the 2012 guidelines for immunizations from the CDC recommend hepatitis A for all patients with chronic liver disease, pretty much as Ray mentioned.

CHAIR BROTMAN: Thank you. Doug, did you have anything to add or Tom did you have anything to add? Your card is up, Tom.

MEMBER CAMPOS-OUTCALT: Yes, I just wanted to comment on that cost-effectiveness thing because being on the ACIP I have sat through a number of these cost-benefit analyses and they have wide confidence intervals, to put it mildly.

But with the rate of hepatitis A that remains in the country even though it is low, it is tons higher than meningococcal meningitis, for instance, which has a rate of 101 per 100,000 and a vaccine that costs a lot more.

So I would suspect that this

probably is going to come in pretty well on a cost-benefit analysis.

CHAIR BROTMAN: Okay, thanks. Are you responding to that?

MEMBER GIORDANO: Differing comment. Oh, go ahead.

CHAIR BROTMAN: Tom, did you want to go first? I'm sorry.

MEMBER MILSTONE: This falls to your point which is if next door they are voting on the measure to use this for chronic hep C or chronic disease, so then I come back to so why then are we going to avoid evidence to have another measure that targets this differently when the reason that you are giving to give it -- so the recommendation to give it to chronic hepatitis people is being discussed next door.

DR. WONG: Yes. That is not my purview but my guess is if it is approved by both groups, which I don't know that that will happen, there will probably some reconciliation process. I will also point out that the level

of measurement is different between the two measures. Theirs is at the system level or the health plan level. This one is at the physician level.

MS. WINKLER: Also I think that we are asking you to look at this measure on its own right now. When we want to look at similar measures, those will be the issues that come in to play. But different levels of analysis are important consideration.

CHAIR BROTMAN: Okay, let's try to wrap this up. Tom?

MEMBER GIORDANO: Just a quick comment. While vaccines are, in general, good and I am a proponent of them and I vaccinate my patients, every time there is a quality measure that is adopted, it means you look more closely at that and often it means you look less closely at something else. There is a shift in what people pay attention to.

So if there isn't strong data to support this rising to the level of an

NQF-endorsed quality indicator, then I think says something.

CHAIR BROTMAN: And I think, Adam, you are going to have the last word.

MEMBER THOMPSON: Yes, I just wanted to add from a patient viewpoint on this one of the things that also I think you have to look at is that the behaviors that lead to hep C infection are the similar behaviors that lead to HIV infection. And there is high comorbidity in that population. And when you are looking at preserving liver functionality, particularly in hep C and then adding HIV on top of it with those medications, I know we are not talking about HIV directly here but the rates of infection in both those populations we are told as patients, and I was previously infected with chronic hep B, we are told preserve that liver, no matter what you do. And that if we don't do that ahead of time, and I think it is a prevention method, and I think on the patient viewpoint, one stick in the arm is worth our

medication working down the road for anything else that might actually kill us.

CHAIR BROTMAN: Thank you for that point. I think we have enough. Let's go to the vote on the empirical evidence, the exception.

MS. WINKLER: Let me just -- this is a vote you haven't taken before. All right? So just so you know how you are voting. If there is no empirical evidence, which is what you have already said, is there an exceptional and compelling reason that the measure should be considered further? In other words, would move on to further evaluation. One is yes, two is no. Any questions about how you are voting and what it means?

MS. KAHN: Okay, you can go ahead and start voting.

(Pause.)

MS. KAHN: So we have 16 yes and four no.

CHAIR BROTMAN: Okay, so it passes.

So now let's go on to scientific -- oh, the gap. I'm sorry. The performance gap.

MEMBER COLLINS: Yes, so the performance gap, the gap is listed at -- the aggregate performance rate is listed at 83.27 percent with a mean of 67.47. And I believe that is the numbers that were listed there. Dr. Wong can maybe comment a little bit further but it appears that there is a gap and if I heard you correctly, those are the people who had incentive to report in the first place.

So I would suspect that it is potentially higher than that.

CHAIR BROTMAN: Dr. Wong, did you want comment at all?

DR. WONG: Yes, again, outside of PQRS, the performance gap is much more substantial, even in the 20 percent gap there.

CHAIR BROTMAN: Any other discussion? All right, let's go to the vote on the performance gap at this point.

MS. KAHN: Voting on 1b,

performance gap; high, moderate, low or insufficient. You can go ahead and start.

(Pause.)

MS. KAHN: You have eight high, 12 moderate, zero low, and zero insufficient evidence.

CHAIR BROTMAN: Okay, great. Let's move on to the scientific rationale -- yes, the reliability portion at this point.

MEMBER COLLINS: As far as reliability goes, the workgroup really didn't have too many comments as far as that goes. High acceptability rate for reliability, I'll perhaps let some other group members comment.

CHAIR BROTMAN: Any of the workgroup members want to make a discussion at this point?

Go ahead, Mohamad.

MEMBER FAKIH: I think if you have EHR, it will be highly reliable the way you can capture that measure. If you don't have EHR, it is going to be tough. That is how I see it.

CHAIR BROTMAN: Does the measure developer want to comment on any of this?

DR. WONG: I think I will defer to anybody.

(Laughter.)

MS. CHRISTENSEN: So I can speak to how we tested that. We had two practices. One a safety net general practice, sort of practice, and the other in more of a specialist practice. Both did have EHRs. Both had been using those EHRs for more than three years. So that is the environment that we tested it in. We ran an automated report out of their EHR, which they built based on our specifications and then did manual chart abstraction to compare the results of the automated report to the manual review and then the reliability was what we presented in our documentation.

So agreed, things are going to be more difficult if you don't have a way to automate the reporting.

CHAIR BROTMAN: Okay Doug, did you

want to say something?

MEMBER CAMPOS-OUTCALT: So how did you capture past immunization records? Because hepatitis A now routine child vaccine, many kids have been vaccinated. Notoriously hard data to get when you are an adult without a vaccine record. How did you get that?

MS. KAHN: Any information that was not in the electronic health record or in the patient's chart is considered not to be real information. If it is not documented, the provider doesn't know about it. So they would, I assume, ask if they had not asked, they would not know whether they should give the patient. Does that make sense?

CHAIR BROTMAN: Aaron, did you have a question?

MEMBER MILSTONE: So I mean I think this is going to come up with other measures because there are a number of vaccine measures. And I will bring the discussion up now and I think it will apply to all, which is how you

capture this evidence of past immunity.

So I think this measure and the next measure actually have a component of capturing laboratory data in the EMR. But in the absence of that, if you have a patient under care for hep C for three or four years or ten years, who had hepatitis A vaccine given ten years ago in a different provider or in a medical record that is now moved to electronic that is not in that electronic record, I think we had some concerns about how often they are missing immunity that is not documented by CPT code or on an active record.

So yes, they have gotten it but the EMR is not capturing it or the physician every year isn't checking the box that is saying yes, this patient had. And that is a big concern I have had for other ones. But recognizing that this will go outside the EMR, I think it is important to discuss it for this one as well and how that impacts validity.

CHAIR BROTMAN: Good discussion

point. Tom, did you want to --

MEMBER GIORDANO: Yes, to follow up on that a little bit, the denominator in 2a1.6, the denominator time window is 12 consecutive months. Does that mean patients who were seen in the practice in the last 12 months are eligible to be included in the denominator or you have got to look at what happened to those patients in the last 12 months. Did they get a hep A vaccine or note that they already had a past vaccine in the past 12 months. Which of those is it?

Because clearly there is no indication for annual vaccination or annual noting that someone is immune. Did you look back in all time whether the vaccine happened and is that what the indicator is asking for?

MS. CHRISTENSEN: So the eligible patients are seen within the 12 months but then any vaccination and I apologize, I am not a clinician, any vaccination that is relevant to the question would then count. Dr. Wong, does

that address the question, do you think? Does that answer your question?

MEMBER GIORDANO: Yes. And are there adequate codes if a person is immune?

DR. WONG: So there is the standard CPT category codes when we see patients. And we all recognize how expensive and how inaccurate paper medical records are and how painful it is to get quality measures from them. And so over the last several years there has been a big push towards electronic health records. But in particular, something you call Category II CPT codes, which provide the opportunity to document these quality improvement, quality assessment measures. And as a quality measure, I think it would provide some incentive for physicians either to ask or to document the antibody level, if they are not sure of the history.

MS. RALLINS: I would like to build on Dr. Wong's comment about documenting immunity.

So in addition to the CPT Category II codes, in our specifications we have been following the recommendations of the HITC Committee from ONC and also including the SNOMED codes that allow you to specifically document immunity, in addition to CPT Category II.

CHAIR BROTMAN: Thank you for the clarification. Mohamad?

MEMBER FAKIH: Just for the developer, it is when we say documented immunity to hepatitis A or hepatitis B, if the provider, if it is not the lab but the provider says or documents electronically, let's say that the patient is immune, does this count as acceptable?

DR. WONG: It depends on the intensity of the investigation. Most of these are designed to be done relatively cheaply through administrative codes. And so if somebody wanted to satisfy quality measure and then went ahead and did a chart review in the electronic health record, yes. But you know,

you have to go through that process on your own.

CHAIR BROTMAN: Adam?

MEMBER THOMPSON: Yes, and I just wanted to bring this up. I agree with the implementation comments as well as the comment down here. I think if it is important enough to give the vaccine, it is important enough to know that it worked as well.

And I know that there was an argument was made that initiation of care is what you are trying to measure but I think if we are looking at getting things closer to the outcome, then it is making sure that the vaccination actually stuck. Because otherwise, I view it like PPDs that are never read and I think cost benefit does come in. If we are just going to be blanket giving it, I think we should make sure that it worked. And is there a reason why you wouldn't have the measure look at the completion of the vaccine versus simply just giving the single dose.

DR. WONG: Yes, we spent a fair bit

of time, not recently, but I was involved in the earlier measures, you know, should we document that they got both hepatitis A shots? Should we document that they got all three hepatitis B shots?

We ended up deciding that the measurement burden associated with that, because there can be a big time gap between the three shots, and they could be with different providers over that period of time in terms of who gave it to them and who is providing the care. And we ended up opting for a simpler quality measure, which again was a lower bar but decreased measurement burden and, at the same time, gave us some indication that the patient was getting at least some benefit and, in particular at least for hepatitis A, around 80 percent of them are going to get antibodies.

CHAIR BROTMAN: Kathleen, you have got the last word. Peter is going to go after.

MEMBER BRADY: Okay, so I just was going to ask a question regarding the

reliability. You report a kappa score of 0.48, which is really on the lower side. It really should be above 0.7. I'm surprised no one has really brought that up and I didn't know if you wanted to comment about that.

MS. CHRISTENSEN: I believe that somewhere in our application, we included our interpretations of the kappa scores. I mean, obviously we would like to see those higher. It does fall within the acceptable standards in the literature.

MS. WINKLER: I actually, I put a slide together that has the kappa values. There it goes. It was in one of your memos but there it is.

MS. CHRISTENSEN: So it is important if folks aren't familiar with kappa, it is different than an agreement percentage, so it is not the same as saying around 50 percent agreement. The interpretation is on the slide here with agreement being slight, fair, moderate. This one is moderate. And the

reason for that is that it is the statistic of agreement beyond chance. So it takes into account that chance agreement, depending on the performance rate on the measure. Kappa and the agreement percentage can vary substantially

DR. WONG: So a kappa of zero would be the 50/50.

CHAIR BROTMAN: Okay, Peter?

MEMBER HAVENS: Thank you. In the numerator it says that you can opt out if you have documented immunity to hepatitis A. Where is that part of the numerator captured in general?

So we have been talking about vaccination. Vaccination is not indicated and people have had natural hepatitis A and so are you suggesting -- is it suggested that testing for hepatitis A be done, shown to be negative, and then vaccination given or just that vaccination be given as a single vaccine and that is assumed?

DR. WONG: So that measure -- this

measure could be satisfied in one of two ways, either of the ways that you mention. We did not want to force testing and then vaccination, nor did we want to discourage testing, if you wanted to see if the patient was positive prior to vaccinating.

In published studied from individuals coming from endemic areas, again a high proportion of them, particularly if they have immigrated from those areas will have hepatitis A antibodies and in those cases, it may be more cost-effective to test for antibody than to just vaccinate. In other cases, say you were born here and raised here, it may be more cost-effective simply to vaccinate.

CHAIR BROTMAN: Okay, and Aaron?

MEMBER MILSTONE: Quick question. Again, because this applies to a couple measures, these vaccine topics, looking at -- and I saw a concern about how we are capturing preexisting immunity. So someone who was vaccinated five years ago or had a test five

or ten years ago. Can someone provide, and this applies to anyone in the room, some guidance as to what proportion of providers and patients that these measures may impact are going to be under electronic medical record in the next two to three years versus on paper record?

Because I think, you know, if we think these are excellent, have great validity and reliability in EMR but they don't using paper medical record, then I am a little concerned about implementing them now versus saying well we have concerns about validity and we are not at EMR yet, you know, we are not in a country that has uniform electronic medical record.

CHAIR BROTMAN: Can anyone address that?

CHAIR SEPTIMUS: I don't have statistics on this but the rate of EMR adoption because of the Affordable Care Act is extremely high. And so I think it is going to be fewer and fewer practices that are not going to have an electronic medical record because they lose

that on incentive dollars because it meets meaningful use. Just as an FYI but I don't have any statistics as to the percent adoption. But I think it is going become pretty commonplace.

CHAIR BROTMAN: And Mary and I think we can go for a vote afterward.

MEMBER BLANK: Could I just get a little bit of clarity on page 21 of this document where it is the population criteria? Because this comes into play for a couple of other measures, too. And I just want to make sure that I am assessing it right. The population criteria section.

So just talking about the -- let's go to the numerator first because which of those "and" statements discusses outside of having an antibody test to it, talks about prior immunity? Is there a way for a physician to know from your past history that you have had hepatitis A without doing an IgG or an IgM?

MS. RALLINS: Excuse me. So can you repeat that question again?

MEMBER BLANK: I'm wondering the numerator statement here. What part of that, just because I am not sure I am understanding correctly, what part of that talks about a prior history besides the laboratory testing? Is there a capability of saying that you have had -- for a physician to draw up a code, CPT II code saying you had it ten years ago?

DR. WONG: So there is a denominator exception which then applies also to the numerator. And the denominator exception would be a medical reason for not administering it. And one of those reasons might be that you have already had an injection. So you wouldn't be in the denominator then, so you couldn't possibly be in the numerator then.

MEMBER BLANK: Just another statement of clarity, if you could just go up a little bit Alexis? The initial patient population, the second bullet that says "and count greater than or equal to two of" does that mean two visits within the measurement time

periods? Is that specified in this measure?

MS. RALLINS: Yes.

CHAIR BROTMAN: Okay at this point, let's vote on reliability.

MS. KAHN: Voting on 2a, reliability; again high, moderate, low, or insufficient evidence. You can go ahead and start.

(Pause.)

MS. KAHN: We have one high, 16 moderate, two low, and one insufficient evidence.

CHAIR BROTMAN: Okay, that passes validity -- reliability, rather. Let's go to validity.

MEMBER COLLINS: Yes, I will be quick with validity. We have already talked about some of the aspects of validity here. One comment or concern was the automated health record's ability to capture exceptions for the measure and kind of how that was done. I was wondering if you guys could explain that a little

bit more. Yes, either one.

MS. CHRISTENSEN: So you can see conveniently we have the logic almost up there. If you want to scroll down just a tiny little bit for me. Of course, it splits over the page.

So you can see that there is a lot of different value sets that are provided for medical reasons, patient reasons they are not given. So to be able to automate this information in the electronic health record, obviously you need to have good electronic health record design and put information in discrete fields using code sets, where applicable.

So an example of a way to do that would be to have a specific place that you would document the refusal of a vaccination or a specific place where you would document that the patient has a documented immunity somewhere else. You know, again, if you are doing stuff in free text, we all know natural language processing may or may not be there but we have

had people be successful setting up their system to capture information discretely.

CHAIR BROTMAN: Any discussion on validity? Let's go to the vote, then.

MS. KAHN: Voting on 2b, validity; high, moderate, lower, or insufficient evidence. You can go ahead and start.

(Pause.)

MS. KAHN: We have one high, 18 moderate, one low, and zero insufficient evidence.

CHAIR BROTMAN: Okay, so that passes. We go to usability at this point.

MEMBER COLLINS: So it should be mentioned again that this measure has been in use since 2008, so I would say it is pretty usable as far as that goes. It has also been proposed for inclusion in CMS's EHR incentive program.

CHAIR BROTMAN: Any discussion points on that? All right, let's go to the vote for usability.

MS. KAHN: Voting on usability;

high, moderate, low, or insufficient. You can go ahead and start.

(Pause.)

MS. KAHN: We have 11 high, nine moderate, zero low, and zero insufficient information.

CHAIR BROTMAN: Thank you. Let's go on to feasibility at this point.

MEMBER COLLINS: And again, I think this measure is very reliable and feasible for implementation.

CHAIR BROTMAN: Any discussion or questions? Okay, let's go for the vote for feasibility.

MS. KAHN: Voting on feasibility; high, moderate, low, or insufficient. You can go ahead and start.

(Pause.)

MS. KAHN: Okay, we are missing one response, if you could all just press it one more time.

(Pause.)

MS. KAHN: We have seven high, 13 moderate, zero low, and zero insufficient.

CHAIR BROTMAN: And finally, let's vote on the suitability for endorsement.

MS. KAHN: And overall suitability for endorsement, does this measure meet NQF criteria for endorsement; yes or no?

(Pause.)

MS. KAHN: We have 19 yes and one no.

CHAIR SEPTIMUS: Okay, now Mohamad, you have the next measure; however -- however --

(Laughter.)

CHAIR SEPTIMUS: I think there is a huge overlap between the B measure and the A measure. So I think probably to help along the discussion, is there anything specific about the hepatitis B measure that would be significantly different from A? I think we can quickly move through all -- we have to go through each section but is there anything in particular

you would like to bring to our attention that would require discussion about hepatitis B vaccine?

MEMBER FAKIH: I am going to be extremely brief. One thing. Hepatitis B is three shots and what they are asking for is only documentation of one shot. And immunity does not happen as good as hepatitis A with one shot. So that is probably the main issue that we need to discuss.

CHAIR SEPTIMUS: That is a great point. John, do you want to explain to us why you chose one shot?

DR. WONG: Here again, it has to do with measurement burden. Typically the three shots would have to occur over a time period. And in fact, if you don't adhere exactly to the zero, one-month, six-month, you can still give three shots. And so because we are doing it over a one-year time frame, you wouldn't get full credit, even though you gave maybe one two of the three shots and the patient didn't show

up for the third.

So again, we didn't want perfection to be the enemy of the good. Thank you.

CHAIR SEPTIMUS: Any comment on this before we start going through all the sections? Because that is the one soft area of this particular measurement. I think Mohamad is absolutely right.

MEMBER FAKIH: You know, I view it as an intent of the healthcare provider to vaccinate the patient and it is a demonstration. So I see it as a positive thing.

CHAIR SEPTIMUS: Tom, and then Adam. Tom?

MEMBER GIORDANO: Just I don't see the workgroup summary in our packet for this measure. Am I missing something?

MS. WINKLER: The separation between the two tables didn't happen.

MEMBER GIORDANO: Oh, okay. There it is. Thank you.

CHAIR SEPTIMUS: We only did it in

your packet, Tom.

MS. KAHN: It is on page 13, for those of you who are looking for it.

CHAIR SEPTIMUS: Adam, do you want to go while Tom is looking for it?

MEMBER THOMPSON: You know just in response to your comment, I think as a patient I would prefer to see providers be comfortable with lower scores and be measuring the complete vaccination than measuring only the single dose. Because gain, I mean with the three, knowing how many people I know, just personally who were vaccinated and then still contracted chronic hepatitis B because the behaviors were so similar.

I mean it just seems to me that you are setting a really low bar. And whereas hepatitis A you are looking at an 80 percent, with hep B the rates are so much lower that it might even the argument could be made for something very different there.

CHAIR SEPTIMUS: There is not only

the initial immune response. Of course as most of you know, the third dose really it gives you sort of an amnestic booster response, which is important in terms of duration of potential protection.

So, Tom did you want to comment again or are you okay? Peter?

MEMBER HAVENS: Since we have somebody from the ACIP, does the CDC recommend testing for an antibody response at the end of successful hepatitis B vaccination series in people with hepatitis C?

MEMBER CAMPOS-OUTCALT: I am not aware of it for hepatitis C. I think the only group that -- there is always caveats on these what if recommendations for vaccines because there are a lot of them. But I think the only group is healthcare workers and workers who are going to be at high risk for hepatitis B. I don't think they recommend testing for antibody HIV but I am not positive.

MEMBER BRADY: It is in the DHS

guidelines for treatment of opportunistic infections that persons with HIV that you check a hepatitis B antibody.

MEMBER HAVENS: Right, so does a similar recommendation exist for people with hepatitis C? Because then the issue of whether or not a single vaccine is an adequate guideline becomes moot, since documentation of hepatitis B surface antibody becomes adequate for a statement of no need for vaccination.

So it would be an alternative to this current measure under consideration and would get around the issue that we are measuring physician intent to do the right thing, instead of actually measuring was the right thing done.

So I just wonder if there is a -- I know the HIV guideline but I don't know the hepatitis C guideline. John might.

DR. WONG: There is no recommendation to measure antibody levels in hepatitis C. There are data, I believe I recall a WHO talk from many years ago where the

statement was there is no recommendation to check antibody levels, outside of other than perhaps healthcare workers because the vast majority of patients had detectible antibodies. So the numbers that come to mind are something like two or three in a million who have developed antibodies if you got all three shots.

MEMBER THOMPSON: Just to jump on that real quick, though, I think that is where hep C advocates like beat the drum around parity. And they would say of course there is one for HIV and there is not one for hep C and I think that is the point the community continues to make.

DR. WONG: I will also just make a distinction between guidelines and performance measures. You know, guidelines are systematic reviews, evidence, benefit versus risk. Performance measures are holding physicians accountable and the issues are do you want to hold them to a stiffer measure, so three shots or perhaps as you propose, the documentation

of an antibody, three shots over what period of time antibody level would cause a lot of testing and would be difficult for both systems and physicians to provide that kind of information.

Again, for us, we didn't want to necessarily penalize physicians who were providing high-quality care, at least at this stage. As EHRs become more mature so that we have a longer track record with immunizations, it will be much easier to do the kind of things that you all are proposing and I would fully endorse that.

But personally, not speaking on behalf of the PCPI, I don't think they are quite there yet with EHRs.

CHAIR BROTMAN: And Mohamad?

MEMBER FAKIH: You know, there is a group that is a non-responder with vaccinations. So if you just look at vaccinating the three shots, you still have about ten percent that will not respond anyway.

So we can't just look at the antibody.

You know, all of them are imperfect, all of these measures. So you know, I don't see a negative looking at one shot. But what I am trying to say the antibody by itself is closer to the outcome but doesn't mean that it is without -- it doesn't mean it is perfect.

CHAIR BROTMAN: Helen?

MS. BURSTIN: I just want to be clear that we are consistent, that we have been saying very clearly that the evidence needs to support the measure focus. So if the measure is a single shot, then I think you have to say the evidence is there. I also think the committee will have to go the path you did in the last measure.

We need to be consistent. We can't be harder on some measures earlier in the day and easier on some later in the day.

CHAIR BROTMAN: Right, we have to look at how it is presented for this presentation.

All right, if there is no other discussion, let's go to voting on impact.

MS. KAHN: Voting on 1a, high impact. Again, it is high, moderate, low, or insufficient. You can go ahead and start.

MEMBER GIORDANO: And that is with a single dose, right?

(Pause.)

MS. KAHN: We have four high, seven moderate, six low, and three insufficient evidence.

CHAIR BROTMAN: Okay, that passes. Let's go on to evidence, at this point. Do you want to present the evidence? I'm sorry, I'm going back to performance.

MEMBER FAKIH: The evidence, you know many studies support the hepatitis B vaccination for hepatitis C patients. A recent study also shows gaps in vaccination. It was a VA population with chronic hepatitis C infection. Also the incidence of superinfection with acute hepatitis B and A in

that study was low but was significantly lower in vaccinated patients. So there is improvement -- I mean there is potential for improvement.

CHAIR BROTMAN: Do you want to make a comment regarding the level of evidence?

MEMBER FAKIH: I'm trying to find -- probably -- I'm looking at what the developer has mentioned, 2a, level C the assent grade.

CHAIR BROTMAN: 2aC?

MEMBER FAKIH: Yes, 2a in level C.

CHAIR BROTMAN: Go ahead with any discussion.

MEMBER COLLINS: So am I right, the level of evidence is very similar with this measure as it was to the previous measure? I think they are very much the same here.

CHAIR BROTMAN: Yes, go ahead.

DR. WONG: If I could just add one thing. I think the evidence for potential harm is actually more substantial because there have been three systematic reviews, albeit not

randomized controlled trials that demonstrate much higher risk of hepatocellular carcinoma when you are coinfected with both hepatitis B and hepatitis C, above the additional effects of one on top of the other.

So and these are larger bodies of patients, as opposed to the hepatitis A.

CHAIR BROTMAN: Ray?

MEMBER CHUNG: And I would amplify that statement by saying that that is only in those patients generally we think as being chronic hep B infected.

So in the adult infection that is 90 percent of patients who cleared, ten percent who may go on chronicity with adult exposure. So that is ten percent infections go on in chronicity and then raising the possibility of a double whammy on that patient for chronic liver disease and cancer.

CHAIR BROTMAN: Aaron.

MEMBER MILSTONE: Well we are not gauging the evidence that hepatitis B worsens

outcomes in patients with hepatitis C. We are gauging the evidence on whether or not one vaccine of hepatitis B may lead to improvement in outcome. Right?

So I guess my question like before, which is is there evidence that one vaccine, which I think why you were saying that evidence is the evidence but we might all agree that this is important to move forward. But I think in terms of evidence --

So I just wanted to make sure we are all clear that we are saying is there evidence that one vaccine of hepatitis B improves outcomes in patients with hepatitis C.

CHAIR BROTMAN: Doug, I'm sorry.

MEMBER CAMPOS-OUTCALT: Yes, this is a process question because I think we are going to -- it seems to me like we are going to go through a rather painful process of voting this down on evidence and then making an exception.

So would it be in order to just move

it, we make an exception right away?

CHAIR SEPTIMUS: You are catching on. I still think we are going to have to vote this down and then go to the exception.

MEMBER CHUNG: Let's go through the criteria and vote it down, if you are going to vote it down.

CHAIR SEPTIMUS: Any more discussion before we get to that point? Raymond, did you want to say anything before? Okay, anybody else with their card up?

Okay, so then let's go to the vote on evidence.

MS. MORGAN: Okay, so I'm sitting in for Adeela. One yes; two no, evidence does not meet guidance for quality, quantity and consistency; and three no, insufficient evidence submitted. You may begin.

(Pause.)

MS. MORGAN: It looks like we are missing two votes. Can you try one more time? Okay, there we go.

So zero for yes; nine for no, evidence does meet guidance; and 11 for no, insufficient information submitted.

CHAIR SEPTIMUS: Okay, before we go to the exception, I just want to let you know I finally found the ICP recommendation for post-vaccine serologies. It is not recommended routinely for adults. I thought that was it but now -- hep B. Yes, it is after three shots.

So now we will go to the exception vote. You want to start?

MS. KAHN: We're voting on the potential exception to empirical evidence.

CHAIR SEPTIMUS: Is there any discussion on this before we vote? Oh, I'm sorry. Go ahead, Tiffany.

MEMBER OSBORN: I guess my question is do we think -- the reason that we want to make this exception, is it because we think that physicians won't do it if we don't have this rule, we don't have this measure?

I'm asking. Sorry. Do we think it

won't happen? Do we think that the vaccine will not be given if we don't have this accountability component that we are providing here?

CHAIR SEPTIMUS: If you look at human nature, the answer is no, they won't give it. And there is a huge gap right now with the current recommendation being what it is. I mean, it is unfortunate but not just in some of the measures we are talking about here but it is in many other aspects of healthcare that unless there is accountability in performance, we don't always voluntarily do it.

CHAIR BROTMAN: The gap may speak for itself.

DR. WONG: And I'll just add that gap was during the Kanwal study was about 20 percent were getting hepatitis A vaccination and about 26 percent were getting hepatitis B vaccination. That is with this measure.

MEMBER BEAL: This is Jeff. I will add this measure might give us some strength in trying to convince payer sources to actually

pay for the vaccine as well.

CHAIR BROTMAN: Doug, -- I'm sorry.

CHAIR SEPTIMUS: I don't want to get into this but just as you know, under the Affordable Care Act, under preventative care it is supposed to be first-dollar covered. So there may be some aspects you may not like but this is one that may encourage prevention.

Go ahead, please.

MEMBER COLLINS: Well you know, on that note, we heard about the hep A vaccine. Is the hep B vaccine, has that been shown to be cost-effective?

DR. WONG: I don't know whether I have looked for that one specifically.

It is cost-effective in non-hepatitis C, so I would, by extrapolation, assume that it is.

CHAIR BROTMAN: All right. At this point, if there is no other -- oh, Mike?

MEMBER FARBER: Yes, I was going to say that to me, the main difference between the

last measure is that one, is that the risk groups are more similar, and two, of the problem of two chronic infections. And I was going to comment that the vote -- that the providers probably won't give it if they won't get reimbursed so that the issue of stimulating reimbursement would be for payers is also an important issue.

CHAIR BROTMAN: And David?

MEMBER SPACH: If we're saying that we are not going to get -- you know, people won't do it because they won't get reimbursed or isn't essentially the stick that is making them doing it, if we are only putting it out on the table that there is one shot that is required, are we then going to be really only putting a stick out there that is to give one dose and we are not going to see three doses giving. So we are really voting this down because we don't like that it is one dose or are we voting it down because we don't like giving hepatitis B vaccine? I think we are voting it down, a lot

of this, because the measures got one dose stipulated.

CHAIR BROTMAN: Speaking of unintended consequences -- Doug, did you have another comment?

MEMBER CAMPOS-OUTCALT: The payment issue, I have no idea whether NQF criteria had to do with payment but on preventative services, if it is recommended by ACIP, it is supposed to be first-dollar coverage in all plans, other than those grandfathered. So I don't think payment is an issue here. All three doses, because it is a three-dose recommendation.

CHAIR BROTMAN: Helen?

MS. BURSTIN: In general, I think it is reasonable to consider cost benefit after you have determined that you have got sufficient evidence and effectiveness. And I think that is what is still a question.

So I was so hesitant last time to answer your question, Aaron because we hadn't

yet established the evidence so it was hard to then invoke cost-benefit.

CHAIR SEPTIMUS: Okay, so seeing no other comments, we will go ahead and vote. Just to remind everyone, the current measure that is under consideration is giving one dose and we are making an exception for that. Okay?

So, let's vote.

MS. KAHN: We are voting on the potential exception to empirical evidence. We are going to vote yes or no. You can go ahead and start.

(Pause.)

(Laughter.)

MS. KAHN: We have ten yes and ten no.

MS. BURSTIN: It's an exception. So exception wouldn't go forward with a tie vote.

CHAIR SEPTIMUS: Okay, so we are going to stop here.

Now tell me, can we give a recommendation to the developer on this issue?

Because if I think I heard, and tell me if I am wrong, if three doses were included in this measure, I think this group would have voted for an exception. Is that correct?

MS. BURSTIN: I don't think you would have had the exception. I mean, the evidence was there, it sounds like.

CHAIR SEPTIMUS: Well, it depends on how you look at it. But the point is if something would have passed, the exception would have passed.

Okay, so let's then go on to 058 -- no, I'm sorry -- 0393. I'm sorry. David.

DR. WONG: Okay so 0393 is a maintenance measure. It was instituted July 31, 2008. I think the question regarding this measure in terms of our group that came up, the biggest issue really was the overall opportunity for improvement. So I will focus some of the discussion on that. Some of this has been addressed by John but I would like to come back

to this.

First of all, just to emphasize what this measure is, it is a measure that is actually specifically looking at -- I'll read the measure here for a second. Sorry, I got off of that.

The measure is the percentage of patients aged 18 and older with a diagnosis of hepatitis C seen for an initial evaluation who had HCV RNA testing ordered or previously performed. And as John has mentioned, the overall importance of hepatitis C I think is really unquestioned right now with approximately three million people living with this disease in the country, approximately four million people having been infected and I would say the perspective on this measure has dramatically changed with the MMWR guidelines that came out approximately ten days ago.

So first of all, to emphasize why this measure is so important, this is the single test that differentiates whether or not a person has been chronically infected with hepatitis

C or whether or not they have resolved infection.

So from the standpoint of clinical importance, it is an absolutely critical measure that determines whether or not a person needs to engage in care for their chronic hepatitis C. So we can vote on that at this point or if you want me to run through, we will run all three of them first. Is that right? Are we going to vote on impact first or discussion on that?

CHAIR BROTMAN: Just impact at this point.

MEMBER SPACH: John, I don't know if you want to add to that or Ray if you want to add to that.

CHAIR BROTMAN: Any comments? Okay, we can go to impact.

MS. KAHN: Voting on high impact; high, moderate, low, or insufficient evidence. You can go ahead and start.

(Pause.)

MS. HAMMERSMITH: We have 16 high, four moderate, zero low, and zero insufficient

evidence.

MEMBER SPACH: Okay, for evidence next, the evidence that was cited in the measure was initially from the ASLD guidelines. This was a category 1b and 1a recommendation. There was subsequently information that was provided by PCPI that included a meta-analysis that included 31 studies and basically these studies all are consistent with an overall estimate of approximately 15 to 20 percent of people who become infected with hepatitis C who clear the virus and thus, this test is important in differentiating whether or not people have resolved infection or chronic infection.

If anybody else wants to make comments on that.

CHAIR BROTMAN: Any discussion? Go ahead, Tom.

MEMBER GIORDANO: I'm not sure I've got this formulated in my head yet but the indicator is getting the HCV viral load measured. And so clearly if you are going to

go down the treatment route, it matters. I mean, it would be hard to show evidence that it -- it may be difficult to show evidence that it matters, but you can't treat someone without knowing that they have viremia. So it makes sense.

If someone is not on the treatment route at all, what is the evidence to say that you need to know whether they are viremic or not, if someone clearly is contraindicated from treatment of Hep C?

MEMBER SPACH: I don't know if I can tell you all the evidence right off hand, Ray may want to comment on this as well, too. But clearly people have chronic hepatitis C, even if they are not on the treatment path right away, they certainly need to be engaged in care where they are getting counseling about alcohol, they are getting counseling about transmission. They are getting information that would be monitoring for cirrhosis and potentially monitoring them for hepatocellular carcinoma.

So I think there are a number of clinical issues that would be relevant.

I can't cite all the data for that and I don't know if John or Ray wants to comment.

MEMBER CHUNG: Yes, I absolutely would again support what David just said, I mean that they have not only branched into a group that should be considered for antiviral therapy but accepting that, that they are not candidates at least for the time-being, these are patients who need to be engaged in long-term care. Staging of the liver disease most importantly because I think with advanced stage disease as is so often the case when we first discover these patients, they need to enter into care for prevention of long-term complications and chronic disease.

MEMBER GIORDANO: I guess my question -- maybe it is not a question. It is just, how do you prove -- what is the evidence to say that that matters and is that in here?

I get what you are saying.

Clinically, of course you need to know if someone has got chronic infection.

MEMBER CHUNG: I mean, this simply, this is algorithmic in a sense. You are really identifying those patients who have chronic infection and, therefore, are at risk for all of the potential complications of the disease. You have to sort them at at least one point in time and then sort them for participation and care and chronic care because that is what they merit, whether it is with therapy or not.

CHAIR BROTMAN: Doug?

MEMBER CAMPOS-OUTCALT: You won't hear me say this very often but I think this is one of those measures that evidence criteria is not appropriate for because it is intuitively obvious. Nobody is going to test it.

Again, you won't hear me say this very often because at my med school I am kind of known for being a hard core evidence person but I think there some instances, not very often and not nearly as often as people advocate for,

but there are some instances where evidence you just can't get and it is not appropriate. And you know, the gray criteria, everybody makes exceptions for this kind of thing. And I didn't see that kind of exception capability in that criteria.

CHAIR BROTMAN: Peter?

MEMBER HAVENS: The results of this test might tell you whom you wanted to offer one hepatitis B or one hepatitis A vaccination.

(Laughter.)

CHAIR BROTMAN: Good answer. Okay, any other discussion?

MS. WINKLER: Yes, just to respond to Doug's question about you didn't see any opportunity for exception. You have just invoked it twice. So you can invoke the exceptions.

MEMBER CAMPOS-OUTCALT: But that is after voting it down based on evidence. I mean, the exception I am looking for is evidence criteria is not appropriate in this instance.

MS. BURSTIN: So just to weigh in on this a little bit because our Evidence Task Force talked a lot about that. And I think one of the issues is at times our assessment measures, which is really essentially what this is, does this patient have this diagnosis, that is the standard of care. The question is, is it a performance measure and do you still need evidence for the measure focus? Does measuring this change the outcome in a way?

So I mean I think it is an interesting question and in fact our Consensus Standards Approval Committee generally doesn't support. I'm curious to see if this goes all the way through what the CSAC will actually say because it is, at some base level, an assessment measure that should be the standard of care.

CHAIR BROTMAN: Any other comments or discussion?

MEMBER SPACH: I would just think the argument that if you can't figure out who is in this group to treat, you are never going

to have any outcome benefit at all. It's like saying prove that testing people for HIV gives them a better -- at some point you have to identify what the disease process is.

CHAIR SEPTIMUS: So let me see if I -- this is one of those deals where you don't need evidence to vote on the measure. Is that what I am hearing?

MEMBER SPACH: Well I guess the question is where downstream we are asking for the evidence.

MEMBER CHUNG: I would say superficially here, you can't have disease without viremia. And to the extent that viremia, that disease equals viremia, then there is your evidence. I mean, it is textbook evidence.

MEMBER SPACH: And there is evidence that more people died in 2007 from hepatitis C than HIV. So if you -- these statistics came out from the CDC so that the death rate in hepatitis C, and you can ask Ray,

has exceeded HIV in the last several years.

So if we are saying that this is a disease that is, as Ray says, if we are identifying viremia and we are identifying that there are people who are dying from the disease, then I think it is indirect evidence but, you know --

CHAIR BROTMAN: Go ahead.

MEMBER GIORDANO: So maybe it is -- I have to keep this in the context of we need evidence to back our decisions up. Clinically, yes, it is obvious you have to know the person has active hep C in order to counsel them, in order to advise treatment and so on. Maybe the evidence that I am looking for is that earlier diagnosis matters. So you want to diagnose. If you find hep C antibody positivity, you need to make sure they have or don't have active replication going on. I don't know but yes, it is a no-brainer on the one hand but on the other hand, how do you prove that this -- that viremia, that knowing whether

someone is viremic or not is better for the patient.

CHAIR BROTMAN: I think the measure developer wanted to respond.

DR. WONG: So there are multiple interventions that are possible in the patients who are detected to be viremic. The one study that I probably would point to is the VA study by Backus, which showed that antiviral therapy in the VA system in patients who have multiple comorbidities, so they could die from many other things aside from hepatitis C where they demonstrated roughly a 50 percent reduction in all-cause mortality related to antiviral treatment, after controlling for multiple confounders.

So this is in addition to Ray's points of sort of 101. In addition some people might consider the alcohol intervention as reducing alcohol intake, but in addition to that, there is the public health benefit of reduced potential transmission.

CHAIR BROTMAN: Doug.

MEMBER CAMPOS-OUTCALT: You know, that study is referred to a lot and what the study, if I recall actually said was that for those patients who showed sustained viral response, that that was a reduction. That makes all kinds of sense. People who are healthier have a sustained viral response are going to have less death. That was not a randomized controlled trial. It wasn't even a study of the treated versus non-treated. It was a study of those who were treated who responded versus those who were treated that didn't. For the life of me, I don't understand why that study has been continued to be referred to as showing evidence of benefit. It doesn't. It is an observational study that doesn't even look at treated versus non-treated. It looks at treated, those who responded, and those who didn't. If I am mistaken on that, correct me but I believe that that is the way that study was interpreted.

CHAIR SEPTIMUS: Well again, I think you've got the criteria for evidence. I don't think we need to go through it again. So I think we are at the point where we need to decide.

Oh, I'm sorry, Peter. I'm sorry, I didn't see you. I apologize but we need to decide on this.

MEMBER HAVENS: Thank you. Listening further to Dr. Giordano, I am swayed by your comments, especially in the context of 0584 which is a timed test of hep C viremia prior to initiation of treatment. And I don't know if we are going to discuss these in the harmonization tomorrow. So we don't bother with that right now?

MEMBER GIORDANO: That's correct.

MEMBER HAVENS: But then I think I share your concern about the timing of the testing and whether or not the initial time is appropriate. I think that is a good question.

CHAIR SEPTIMUS: Kathleen.

MEMBER BRADY: I just wanted to make a comment about the fact that this is -- I don't see this as a performance indicator. I see it as a diagnostic algorithm and that, I mean, no different than the way we diagnose the new HIV testing algorithms where you would do an EI, fourth generation EIA followed by a multispot or a NAT. I just don't see this any differently than that kind of situation where you are trying to figure out who has disease and who doesn't. And I'm not sure that that really should be a performance indicator.

CHAIR SEPTIMUS: Yes? Doug, do you want to do it?

Just to complicate it there is maybe another situation that has not been mentioned where detecting viremia earlier in acute disease and treatment does clearly influence outcomes but that is a very specialized situation.

So you ready to vote? I guess we are. Let's vote.

MS. KAHN: Voting on 1c, evidence.

We will vote 1, yes, the body of evidence meets the guidance for quantity, quality and consistency; 2, no, the evidence does not meet the guidance for quality, quantity and consistency; or 3, no, insufficient information was submitted to rate for quantity, quality and consistency.

You can go ahead and start.

(Pause.)

MS. KAHN: I think we are still waiting on two people.

Okay, everyone just push it one more time.

(Pause.)

MS. KAHN: So we have three yes, the body of evidence meets the guidance; eight no, the evidence does not meet the guidance; and nine no, there is not sufficient information submitted.

CHAIR SEPTIMUS: Okay, well I think you know this one fails. We could ask the question should we make this an exception or

not, as we did for the other two measures or do you think it just fails altogether?

Kathleen, I know -- so is there anybody who wants to vote on an exception for this measure? Then that ends the discussion on this measure.

Okay now before we can take a break so we can get to the 1:15 mark --

(Laughter.)

CHAIR SEPTIMUS: Is it safe to walk back to the hotel after dark?

(Laughter.)

CHAIR SEPTIMUS: All right, 0584.

CHAIR BROTMAN: That's me.

CHAIR SEPTIMUS: And that's our co-chair.

MS. WINKLER: But we change major developer at this point. So Dr. Clyman is he -- where did he go? He's right over here.

CHAIR SEPTIMUS: And thanks to Dr. Wong who really did a wonderful job. Thank you.

DR. CLYMAN: Thank you. My name is Jeff Clyman. I represent Resolution Health.

Measure 0584 looks for a quantitative RNA measurement within a six-month period preceding the initiation of pegylated interferon therapy as treatment of chronic hepatitis C infection.

The primary issues in question today concern the overlap with measure 0395 developed by the PCPI. While the two measures have the same intent exactly, they are optimized with distinctly different information and sources. The PCPI measure appears to be geared toward interoperability with electronic health records, focusing on the data pertaining to an individual provider's practice.

In contrast, measure 0584 relies upon an administrative data set which is typically available to health plans and insurance companies and is likely to represent a broad picture of a patient's healthcare experience, extending well beyond the

contributions of a single provider.

As such, the measure closely follows the formulation of traditional HEDIS quality measures. For example, by requiring sustained period of continuous eligibility for both medical and pharmacy benefits. These constraints significantly enhance accuracy by assuring the presence in the data set of a billion claims for all rendered services, enabling correct conclusions about the initiation of drug therapy and the absence of a viral load test.

Several additional characteristics of measure 0584 further underscore important differences with the PCPI measure, again reflecting alternative perspectives. And I am happy to enumerate them as appropriate.

Thank you.

CHAIR BROTMAN: Okay, I am going to go through this but as you heard, there is going to be a harmonization issue that creeps up with this with 0395 relating to the type of source

of claims and so forth.

But let me go through this. The measure itself is hepatitis C viral load test, 0584. The description reads that: "This measure identifies the percentage of patients with chronic Hepatitis C who began HCV antiviral therapy during the measurement year and had HCV Viral Load testing six months prior to initiation of antiviral therapy."

It is at the level of analysis at the health plan level. It is a process measure, maintenance measure originally endorsed in 2009 and based on the source of administrative claims, as you heard.

I will just go through impact quickly and then we could probably vote on that. There is currently we have talked about the importance of how hepatitis C has been a major disease burden in the United States and the testing was important. Prior to starting therapy for multiple reasons, additional notations by this measure developer state that

the viral load prior to treatment is critical for assessing virologic response during antiviral therapy to tailor treatment duration, including shorten treatment course and termination due to fertility, that is unlikely to become viral negative with prolonged antiviral therapy.

So given that, we could probably vote on impact, unless there is any discussion. Okay, so let's vote on impact.

MS. KAHN: Voting on 1a, high impact; high, moderate, low, or insufficient evidence.

(Pause.)

MS. KAHN: If we could have everyone try one more time.

(Pause.)

MS. KAHN: So we have 11 high; six moderate; one low; and one insufficient evidence.

CHAIR BROTMAN: Okay, so that passes. Let me talk a little bit about the

evidence at this point. Evidence is based on a clinical trial guideline, which reports the level of evidence in a Class I, Level A which was assigned by the American Association of the Study of Liver Diseases, which based it on the American College of Cardiology and American Heart Association Practice Guidelines. And specifically, there were 12 clinical trials that were studied in the meta-analysis paper and the studies themselves followed.

The quality of evidence they followed in the number of patients ranging from 70 to 731, there were similar results speaking to consistency across the meta-analysis, showing that obtaining a base viral load of HCV patients is beneficial. And so there appears to be quality, quantity and consistency addressed within this guidelines presentation.

Any discussion? Yes, go ahead, Peter.

MEMBER HAVENS: Are we to evaluate this without the clarification that we would

be identifying which HCV type it is, since knowing the type is crucial to treatment and this is a test that would be done prior to treatment? This is just a virus measurement, without identifying whether it is one, two, or three. Is that -- I'm just trying to --

CHAIR BROTMAN: I believe that is correct.

MEMBER HAVENS: The next series talks about harmonizing all the -- again, I am trying to understand. Because in clinical practice, you need to know what type it is to make a rational treatment decision. And so this kinds of gets to Tom's prior --

DR. CLYMAN: This measure is not meant to imply that --

CHAIR BROTMAN: You have to put your mike on.

DR. CLYMAN: Yes, this measure is not meant to imply that the only prerequisite to beginning drug therapy is the viral load test. This measure simply looks for the performance

of the viral load test, you know, understanding that there may be other things that are necessary before commencing therapy.

MEMBER HAVENS: Okay.

CHAIR BROTMAN: Yes, Raymond?

MEMBER CHUNG: So I wonder if this gets us into the issue of bundling for hep C preparation for therapy. I mean, I know the following two are quasi-bundled, 0394 and 0395 was it? Whatever. Genotype plus viral load, they were two consecutive items. But I wonder if that is kind of where we are headed with all of this and whether at the end of the day a conference committee putting this together into some kind of unified hole.

DR. CLYMAN: Well, NQF, I think we will try to harmonize these things.

MS. WINKLER: I think that we definitely want to do that but we are talking about those two measures are clinician level measures. This is a health plan level measures. So we do have those differences. The question

I would pose back to resolution health is have you considered measuring things like a genotype measure prior to therapy so that your measure might be more comprehensive about pre-therapy evaluation.

DR. CLYMAN: It is something we are looking at. In fact, I believe we do have a measure that looks for performance of the genotype measurement. It is just not included in this.

MEMBER HAVENS: So before you start therapy, you want to identify that somebody truly had chronic infection. So you need to do a viral load test. But if you are really going to think about starting therapy, you need to know the genotype so you can make appropriate plans for therapy and follow-up.

So if this is a pre-therapy test, standing alone, it seems inadequate. I am glad to have you tell me why that is wrong.

DR. CLYMAN: Well, I would not suggest that that is wrong. I would consider

this to be an individual measure and a possible composite measure would be one that combines the two individual measures, one looking for a genotype measurement and the other for viral load measurement.

We chose to only address one of those measures presently.

CHAIR BROTMAN: Raymond?

MEMBER CHUNG: Maybe I am peering too far into the future but just in response to that question, I would say that we are headed toward a world that will become genotype independent from the vantage point of selection of therapy. We are not there yet. And in fact, genotype, as we will talk about later has more to do with duration perhaps as much as -- and with such protease inhibitors.

But I would say that ultimately we hope with the pan-genotypic therapy, the viral load will be the only thing that we really have to care about prior to initiating treatment.

So I think evolutionarily speaking,

this will stand alone. And so how we want to handle that is, I suppose, technical. But I think it can be considered on its own merit for the time being.

CHAIR SEPTIMUS: Kathleen?

MEMBER BRADY: Well I'm certainly not a hepatitis C expert but what I wanted to ask is, I mean, in the denominator statement it requires a new start of peginterferon in the last year. And although I think all the current regimens most people are still using peginterferon, I think the future that is not necessarily going to be the case.

So do we want to commit to a measure of using a drug that could quickly become outdated?

CHAIR SEPTIMUS: Adam?

MEMBER THOMPSON: Yes, I just wanted to ask and correct me if I am wrong about this, but I think that the difference I see here is that this isn't about identifying what genotype you have, which would happen at the

beginning like diagnosis. This could be someone who has already had the genotype test but didn't have a viral load. And then as they progress and the disease may need treatment later. So you are just ensuring they have a viral load before they are treated. But the genotype may already be known. Correct?

DR. CLYMAN: Yes.

CHAIR SEPTIMUS: Tom.

MEMBER GIORDANO: So these are people -- to be in this measure you have to already -- you get treatment, right? And these are people who are going to get or who have gotten treatment because the denominator is a retrospective look back at people who got treatment. Did they have a viral load within the six months pre-treatment?

So is there any situation where that wouldn't apply? Like if someone was known to be viremic a year ago and you knew they got or you thought they get HCV from injecting drug ten years ago, would you necessarily need to

repeat that viral load prior to treatment, if you knew that they were viremic already but it was more than six months? Maybe that is too detailed at this point in the discussion.

MEMBER CHUNG: Yes. Sorry. I'm not sure where the six months actually came from. That is kind of a number drawn out of a hat because you can make the argument that could be 12 months, that could be 18 months. The mere point is that you actually want a viral load that is sufficiently proximate to the start of therapy, normally because you want to document viremia but just as importantly you want to know what the magnitude of the viremia is as they start treatment. And that is really, I think, the point behind the six months, you know, that the magnitude of reduction of the viral load does matter during therapy.

MEMBER GIORDANO: Does it change in the natural history of the disease? Does it change like HIV?

MEMBER CHUNG: Well it can. It

can. I mean, certainly there are fluctuations by half a log or so over the course of a chronic infection. And certainly in latent disease, viral loads may even drop with advancing cirrhosis.

But I think to the point of where log reductions matter during therapy for stopping rules for treatment, you want to have as accurate a barometer of where they were just before therapy.

CHAIR SEPTIMUS: Let me see if I can summarize where I think the comments are going.

If this measure had hep C viral load with a genotype with an exemption if the patient already had a known genotype, how would that measure be if it was proposed to the committee?

It seems like the hurdle here is that you should know the genotype before you start treating but this measure doesn't require it and there may be some patients where the genotype may be known, where repeating the genotype would be redundant in excess cost. So could that be

an exception if you already knew the genotype?

DR. CLYMAN: Well I think Dr. Chung points to an interesting scenario that the measure, that the single, the individual measure addresses and that is lead treatment. You don't need to repeat the genotype every time you commence a course but the recommendation is to obtain a baseline viral load before repeating every course of therapy.

CHAIR SEPTIMUS: Maybe I misunderstood the measure. Does it say for repeat treatment?

DR. CLYMAN: No, it is for treatment.

CHAIR SEPTIMUS: Treatment only, so it would cover both then, would it not?

MEMBER CHUNG: Yes. I guess we are kind of getting bogged down in the grouping of genotype with viral load. I think it is a true statement that you need a viral load before therapy. And I think that is what this statement addresses. That you need another

test is, I suppose addressed in a separate statement and I guess we will get to that.

MEMBER CAMPOS-OUTCALT: Given how this is so accepted and is really the standard of care, and this comes up in a couple of measures that we will look at after the break.

Same question. I find it so hard to believe that this isn't being done at a pretty high rate already.

DR. CLYMAN: Yes. In our analysis of more than 1.5 million members in a commercial insured populations, we found that the compliance with this recommendation was roughly between 70 and to 85-90 percent. So there is significant opportunity.

CHAIR BROTMAN: So let's just stay on the evidence. Raymond?

MEMBER CHUNG: Do you know if the failure in that gap was related to having gotten the viral load many years earlier or having not been in the six month window? You know, you described a 25 percent failure rate or a 20

percent failure rate there. You don't know.

DR. CLYMAN: No, I do not know the reason but again the measure looks for a new start of therapy. This is not the first course of therapy. This is just a new start of a course of therapy.

MEMBER CHUNG: You know, this doesn't square so much with is just a real world experience where our insurers and our third-party payers ask us what the viral load was, as a pre-condition of a prior authorization. That is the funny thing in all of this that such a gap does exist. I mean, it is a little bit, you know, sort of a disconnect.

CHAIR BROTMAN: So we are continuing to talk --

MS. WINKLER: I just wanted to respond to Kathleen's comments about evolving, changing therapies and new drugs coming along and new regimens coming along. This is not unique to this measure or this topic area by

any means. We see this all of the time.

So one of the things that happens is that these measures are not static, they are dynamic. It is the reason we do check in with the developers on an annual basis. There are likely to be updates as new drugs, new regimens, new recommendations come along and the measure can live along with it.

So even though you can project changes in the future, you don't need to do that right now. That is kind of part and parcel of how we will carry the measure forward.

CHAIR BROTMAN: Any more discussion? We are talking about the evidence, the quantity, quality and consistency of it.

Doug, did you have your card up for a reason? Okay, so let's go for a vote on the evidence at this point.

MS. KAHN: Voting on 1c, evidence. So one, yes, the body of evidence meets the guidance; two, no, the evidence does not meet the guidance; and three, no, insufficient

evidence was submitted. You can go ahead and start.

(Pause.)

MS. KAHN: We are missing two people.

(Pause.)

MS. KAHN: You can just keep pressing until it turns 20.

So ten, the body of evidence meets the guidance; five, no, the evidence does not meet the guidance; and five, no, there is insufficient information submitted. So it is tied.

CHAIR SEPTIMUS: We have to -- it's a tie.

MS. WINKLER: I guess I would like to ask the people who are saying that the evidence does not meet for quantity, quality and consistency, if you could perhaps explain that. I mean, do you really feel that there aren't several studies of good quality showing consistent results that you should do a viral

load prior to beginning treatment? I mean, is that what your no vote means? I'm trying to grasp that.

I have to try and interpret what you said.

And for the other five that voted no, do you feel there is insufficient information provided here to know what the quality, quantity and consistency is?

CHAIR SEPTIMUS: Aaron?

MEMBER MILSTONE: I just want to agree with you because I think people are going to have to say the same thing for the next measure. So it has to be the evidence that you are voting on, not the fact that there is something else about the measure you don't like because then we have to be consistent with the next one.

CHAIR SEPTIMUS: Mohamad?

MEMBER FAKIH: I think it is the timing that strikes me. So if it is seven months versus six months, you know, why would it be

six months before -- within six months? That is what worried me.

CHAIR SEPTIMUS: Anybody else who voted the last two categories who would like to express their reasons? I mean, we are a friendly group. If you survived this morning, you can survive this afternoon.

MEMBER MILSTONE: I just want to add I just looked. That six month window is also in the next measure as well.

MEMBER CHUNG: I am not voting. I am not explaining a no vote. I am simply asking whether we ought to just word that as within six months. You know, just as a -- I don't know if six people are interpreting that literally as six months prior to.

But within six months? Okay. Okay, fine. I don't think that should be a sticking point, honestly.

MEMBER GIORDANO: I agree with that. If we could vote again, if that is the stipulation that it is just within six months.

CHAIR SEPTIMUS: So there was some confusion about that?

MEMBER CHUNG: Yes, there was confusion about that.

DR. CLYMAN: That certainly is the intent and my recollection is this was an area that we deliberately harmonized with the PCPI measure. And there is no evidence surrounding the exact number of days preceding start of therapy that the baseline viral load needs to be performed. We thought that the six month period built into the PCPI measure was reasonable.

CHAIR SEPTIMUS: Doug was first.

MEMBER CAMPOS-OUTCALT: Yes, this is a process question. This is kind of the first vote we have taken that the vote was questioned. And I am kind of wondering why. Is it because it is a tie? Okay.

MS. WINKLER: We are trying to figure out what to do with the tie and what it really means.

CHAIR SEPTIMUS: Thomas? Is your mike on?

MEMBER FILE: You probably don't want to hear what I say anyway. But I mean you just said there is no evidence for that six months, right? You just said that.

So how can you support that third option? The third option says there is insufficient evidence.

DR. CLYMAN: Well I think the question of the exact length of the time interval preceding the first does of the drug is not clear.

CHAIR SEPTIMUS: Adam?

MEMBER THOMPSON: Yes, I just wanted to add when I read the sixth month part and this is just me thinking as a patient, the way I thought about it was that is generally the frequency at which we see hepatologists. We will go in and see them and I wouldn't want someone treating me that hadn't seen me in the past six months.

So I looked at it not as a scientific thing but as an indicator that I was in care and seeing my physician.

CHAIR SEPTIMUS: Yes?

MEMBER CHUNG: Could I propose a revote?

CHAIR SEPTIMUS: Yes, Raymond?

MEMBER CHUNG: Could I propose a revote?

CHAIR SEPTIMUS: Tiffany has a comment, too.

MEMBER OSBORN: I just wanted to clarify the 12 studies that were in the meta-analysis, those were observational or can you -- I just don't remember.

DR. CLYMAN: I'm not certain.

MEMBER OSBORN: So we don't know if they are observational randomized controlled trials. We don't know about the data in the meta-analysis?

DR. CLYMAN: I honestly don't.

CHAIR SEPTIMUS: Okay, so Raymond

proposed that we take a revote. You are not ready, Rekha?

MEMBER MURTHY: It is just to clarify part of that. So if we can just look at the beginning of this measure information, can we just agree just have a consensus that there is basically an error or like incomplete wording that both, under the description and numerator statement the word within six months prior to initiation is what was intended before we do the revote?

CHAIR SEPTIMUS: That's my understanding.

Okay, so we have a motion by Raymond to revote. Is there a second to that?

(Show of hands.)

CHAIR SEPTIMUS: All those in favor, to see if you are awake say aye.

(Chorus of ayes.)

CHAIR SEPTIMUS: Okay, we will revote.

MS. KAHN: So voting again on 1c,

evidence. Yes, the body of evidence meets the guidance; no, the evidence does not meet the guidance; or no, there is insufficient information submitted.

You can go ahead and start.

(Pause.)

MS. KAHN: We are missing one person.

So we have 13, yes, the body of evidence meets the guidance; two, no, the evidence does not meet the guidance; and five, no, insufficient information was submitted.

CHAIR SEPTIMUS: So we had a bunch of flip floppers on number two. Okay, let's keep going.

CHAIR BROTMAN: Okay, we are going to address the performance gap just briefly. And as John mentioned before, the modified measure was tested on three data bases, approximately 1.8 million administrative claims totally. And the data bases consisted of 410,000 claims, 700,000 claims, and 700,000

claims respectively. The results varied from about 70 to 85 percent, as he mentioned with clients. So there appears to be a fair enough range for performance improvement.

CHAIR SEPTIMUS: Any discussion on this one? I thought we'd get by. Go ahead.

MEMBER CAMPOS-OUTCALT: I just have to be honest and say that that raises questions in my mind regarding the reliability and validity of that test. I just can't believe that that is the current statistic.

CHAIR SEPTIMUS: Any other comments? Okay, lets vote on the performance gap.

MS. KAHN: We are voting on 1b, performance gap; high, moderate, low, or insufficient evidence. You can start.

(Pause.)

MS. KAHN: One more person. We have four high, 14 moderate, two low, and zero insufficient evidence.

CHAIR SEPTIMUS: Okay, well that

passes. And so reliability, Mr. Co-Chair.

CHAIR BROTMAN: Let's see reliability. It is mentioned by the measure developer pretty much the same statistics that the measure identified members correctly on three databases. The compliance range from 70 to about 85 percent over 1.8 million claims.

Yes, that is all I have to say I think for that.

CHAIR SEPTIMUS: Comments on reliability? Okay, we'll vote. No. Sorry. Tom.

MEMBER FILE: Does that really apply for the reliability of the data? I mean, is it reproducible? Does it specify specifically what the measure is intended to?

CHAIR BROTMAN: Yes, this was the information supplied by the measure developer.

MEMBER FILE: So you have measures of testing.

CHAIR BROTMAN: Can anyone speak to that? Developer?

DR. CLYMAN: No, we can't.

CHAIR BROTMAN: Okay.

CHAIR SEPTIMUS: Yes?

MEMBER HAVENS: Specifically in 2a2.3 testing results it states what the compliance was but it does not state that any test of the reliability of the measurement as designed.

So there is no reliability measure that I can see in this document, unless I am missing something.

I'm on page ten, 2a2.3 testing results.

DR. CLYMAN: That's correct.

MEMBER HAVENS: It states the compliance varies from 68 to 84 percent but that is not a measure of the reliability of the -- it is not the estimate of the reliability of the measure at hand.

MEMBER CAMPOS-OUTCALT: Right.

MEMBER HAVENS: Thank you.

CHAIR BROTMAN: And that is the only

information supplied by the measure developer for this section.

CHAIR SEPTIMUS: Any other comments about reliability? Then I guess we should vote on this.

MS. KAHN: Voting on 2a, reliability; high, moderate, low, or insufficient evidence. You can go ahead and start.

CHAIR SEPTIMUS: I am told perhaps if we would click our clickers towards the computer, that maybe that will help. Right there.

MS. KAHN: You got it. We have one high, five moderate, four low, and ten insufficient evidence. So it will not go forward.

CHAIR SEPTIMUS: This is one of the stop measures. So this means that this measure fails.

We are going to take a ten-minute break and we will restart. And just remember,

it is just 1:15.

(Whereupon, the above-entitled matter went off the record at 3:29 p.m. and resumed at 3:38 p.m.)

CHAIR SEPTIMUS: We need to have a public comment on the previous measures. So can you open up the lines and see if anybody has public comments from the previous discussion?

OPERATOR: If you have a comment, press *1 on your telephone keypad.

(Pause.)

OPERATOR: And at this time, there are no comments.

CHAIR SEPTIMUS: Thank you very much.

The next set of measures 0395 and 0396, 0397, and 0398 and I guess 0394 and 0401 are all from the AMA-PCPI, which Dr. Wong is back.

So why don't we maybe succinctly to comment on one measure at a time. Let's start

on 0395.

John, are you ready?

DR. WONG: I'm sorry, I was just waiting for you guys.

CHAIR SEPTIMUS: I'm sorry, 0395 and then I think the next two measures will be Doug's. So you get a chance to speak.

MEMBER CAMPOS-OUTCALT: Is he going to comment?

CHAIR SEPTIMUS: Yes, he's going to comment first.

DR. WONG: Oh, I didn't realize that. I thought you were going to start.

Well, --

CHAIR SEPTIMUS: We'll just do one at a time, John. You may have already said some of the things you needed to say.

DR. WONG: Yes, I think so. This is testing for your viral load before initiating treatment. Multiple reasons to do so. One is to document viremia so you could avoid unnecessary treatment of those who are viral

negative and secondly to evaluate viral response to therapy, which are critical for some of the stopping rules.

CHAIR SEPTIMUS: And then just for the clarity, 0396 there is a -- this may be very well paired with 0396.

DR. WONG: In this case, we have elected to pair it with genotype testing because obviously your genotype affects both treatment and treatment duration.

CHAIR SEPTIMUS: So before I turn it over, is this -- are we going to consider each of these measures separately by the standards? Okay. Okay, so we are going to just -- why don't you start with 0395?

MEMBER CAMPOS-OUTCALT: All right, well this measure is the percentage of patients aged 18 years and older with a diagnosis of chronic hepatitis C who are receiving antiviral treatment for whom quantitative HCV RNA testing was performed within six months prior to initiation of antiviral treatment.

Regarding the importance, the data that was presented had to do with a proportion of or the prevalence of hepatitis C, the morbidity and mortality related to that. And when we discussed this as a workgroup, really the only question we had was on the scientific data, what we were presented with was a guideline. And the guideline, according to the assessment of it, did not actually grade the evidence or talk about contradictory evidence and didn't rank it. So I would be interested in comments on that.

And then the performance measure, that was conducted came up with a high 80 percent performance already. So we had a question about what kind of impact we were going to have by adopting this measure.

CHAIR SEPTIMUS: Okay, so of course the first thing we consider is impact. So any comments on impact or would John, do you want to comment on that?

DR. WONG: Just in terms of the

evidence I provided had to do with patients who had spontaneously become viral negative and it does occur in the literature.

The data that I did not provide but is substantial is based on randomized controlled trial data where in all of the registration trial, the viral load was measured at week zero and then assessment was done at regular intervals from four weeks out to 48 weeks.

In terms of the gap or performance gap, again I will mention that the PQRS data are from mostly physicians who volunteer and who will get an incentive in pay and, thus, are incented to adhere to the performance measures.

In the *Annals of Internal Medicine* paper by Kanwal, surprisingly only about 60 percent of patients had a baseline viral load done within the prior six months.

CHAIR SEPTIMUS: Okay, any other questions? We will be talking about the impact first and then we will get to evidence. So any other comments about impact?

Seeing none, I guess we will vote.

MS. KAHN: So we are voting on 1a, high impact; high, moderate, low, or insufficient evidence. Go ahead and start.

(Pause.)

MS. KAHN: We have nine high, ten moderate, zero low, and zero insufficient evidence.

CHAIR SEPTIMUS: Okay, so we certainly passed the high impact. So let's now go to the evidence.

MEMBER CAMPOS-OUTCALT: Well as I stated before, the evidence that was presented to us to discuss was this guideline. And we didn't have a lot to go on. So I think we are going to be more dependent on what is presented here than what we had presented to us in the presentation or in our discussion.

DR. WONG: Just to reiterate what I said before so I won't go through that again, I would just add a side comment that the PCPI in the past have relied extensively on the

guidelines in allowing the guideline process to do the evidence review. At least in the past they have relied on the level of evidence and the strength of the evidence in terms of conveying these things.

As such, because of the request from this particular group, we have decided to go ahead and supplement those data, in particular because some of the criteria you are being asked to evaluate these recommendations on specifically one of the attributes is evidence. So I can understand your need to have that kind of information.

CHAIR SEPTIMUS: Did you say you supplemented it?

DR. WONG: The document that was sent on Monday.

CHAIR SEPTIMUS: I got you.

DR. WONG: And then anything orally I provide that wasn't in the written.

CHAIR SEPTIMUS: Right, okay. So you should have gotten something on Monday about

this. And that is what I wanted clarification. Thank you.

CHAIR BROTMAN: So that is the paragraph that says in 111 patients with biopsy-proven hepatitis C followed for more than five years, two patients spontaneously resolve their infections without any antiviral treatment. In 1667 patients with a history of injection drug use with hepatitis C infection assumed to be chronic, 90 out of 919 cleared the hepatitis C virus over 85 months.

CHAIR SEPTIMUS: Any other comments from you, Doug?

MEMBER CAMPOS-OUTCALT: No, I think that is what we have.

DR. WONG: And I would just add the RCT data where they checked baseline viral load to assess stopping criteria for futility where you need to know the baseline and then you need to know the viral load to climb. And if you don't meet those, then you may discontinue therapy.

CHAIR SEPTIMUS: Okay, seeing no comments, I guess we can vote on the level of evidence.

MS. KAHN: Voting on 1c, yes, the body of evidence meets the guidance; no, the evidence does not meet the guidance; or no, there is insufficient information submitted.

So you can go ahead and vote.

(Pause.)

MS. KAHN: We have 13 yes, the body of evidence meets the guidance; two no, the evidence does not meet the guidance; and five there is insufficient information.

CHAIR SEPTIMUS: Okay, the next is going to be opportunity.

MEMBER CAMPOS-OUTCALT: This next section is the opportunity for improvement. So the data we had presented showed that there was this higher rate of adherence already high 80 percent. So we did not have evidence presented to us that it was lower. So we voted based on that. That was probably the biggest

question mark we had as a group.

DR. WONG: Again, I would just remind you PQRS is a selected subset of patients -- of physicians. It is about 24 percent who have opted into this performance measure and as such, they get compensated if they perform to that measure.

So I would put to you that it is likely a self-selected group who is most likely to adhere to these.

In Kanwal's study involving 14 million patients, only about 60 percent of patients had a baseline viral load tested.

CHAIR SEPTIMUS: Mohamad?

MEMBER FAKIH: Within the United States, do we have data? You know, he said 14 million. I am assuming this is out of the country.

DR. WONG: Those are U.S. patients.

MEMBER FAKIH: These are U.S. patients? Okay.

DR. WONG: So these are 14 million

members in the insured. Not all of them had hepatitis C, just to be clear.

So in this database of 14 million individuals, among the group that had hepatitis C and got treated, 60 percent of them roughly had a viral load prior to treatment.

CHAIR SEPTIMUS: We have a comment from the peanut gallery behind me.

MS. CHRISTENSEN: So after we submitted, we got the 2010 data from PQRS. This is 2009 data. More people are reporting the performance rate has dropped. It is now 23.05 percent on average for the 2010, reflecting more people reporting. So that is a significant difference.

CHAIR SEPTIMUS: Go ahead.

MEMBER HAVENS: Given all of the information about the importance of measuring virus load to A, identify the diagnosis, and B, to make treatment decisions, then what your statement makes me believe is that the way you are measuring whatever it is you are measuring

is not capturing what you wish you were capturing.

So given everything that has been discussed here today for your compliance with this measure to nominally drop from 60 percent to 20 percent, it doesn't suggest to me that physicians are doing things worse as they treat more patients but rather you are not capturing what you want to capture.

DR. WONG: So that is not our measure. We have the measure but the measurement is being done by CMS in their PQRS population. And it is their report. We are not -- correct me if I am wrong. We are not involved with how they measure it, in whom they measure it. It is a population of physicians who volunteer to have themselves measured again. And in that population, that is what is being observed.

MS. CHRISTENSEN: And if I can just add, this is the first one that has changed more than two percent, which is why I have not brought

up the new data before. But we do tend to see that in the PQRS program as more people come on and report in new years, they are not doing as well. They find out how they are doing and then ostensibly, they probably do some quality improvement and start doing better.

MEMBER HAVENS: So you think that is an accurate measure of practice and not a problem with the reliability or validity of the measurement process itself.

MS. CHRISTENSEN: Well when we tested the reliability and validity, they came out very well. So I think the measure is reliable and valid. We helped talk with CMS about their results and they feel that their results do not need to be audited because reporting incorrect information would be fraud and abuse in the program. So they feel that their data is accurate and that is all we can really do there.

CHAIR SEPTIMUS: Raymond?

MEMBER CHUNG: I'll admit -- I'll

accept some data as being reasonable evidence that there might be -- there could be a gap. But 23 percent is I think a little bit -- makes me incredulous. You know again, if this is right that antiviral therapy was administered during a given time period and there was no RNA check during the six months preceding it, is a little beyond the pale. Only because, again we talked about this just logistically, it is very difficult to get away with that and get a patient a prescription, honestly.

DR. WONG: So this is a couple of things. One is, this is a matter of care. Right, so it is not an HMO who is monitoring you or making you jump through a hoop to prescribe the medications. Secondly, in all likelihood the denominator has changed because people realize there is no money in pay for performance. So they may sign up but not fully realize the full set of performance measures. So you may have a whole bunch of folks who are signing up and who potentially are getting

treated, I'm not sure by who, without confirming viral positivity.

CHAIR SEPTIMUS: The only other explanation I can think of is attribution where people are trying to get credit for the measure but somebody else is treating them. And that is the only thing that -- this is really a strange one.

Now, is the measure, refresh my memory, for PQRI I know it is obviously is a type II code. Is it a pay for reporting or a pay for performance?

DR. WONG: Pay for performance.

CHAIR SEPTIMUS: It is now a pay for performance?

DR. WONG: Oh, reporting. Sorry.

CHAIR SEPTIMUS: Reporting. So that is slightly different also. That is what I thought. I think they want to turn it into a pay for performance once they get baseline data but that is the only other explanation I can think of why the numbers change so

dramatically.

But is sound like either way, whether you believe the original data or the new data, it sounds like there is an opportunity, it sounds like. So anyone else want to comment before we vote on the performance gap?

Okay, then let's vote.

MS. KAHN: Voting on 1b, performance gap. Again, it is high, moderate, low, or insufficient evidence. You can go ahead and start.

(Pause.)

MS. KAHN: So five high, 14 moderate, zero low, and one insufficient evidence.

CHAIR SEPTIMUS: Okay, we are going on to reliability.

MEMBER CAMPOS-OUTCALT: Well prior to the conversation we just had, we felt pretty good about the reliability. The data we were presented to consider the test measurement looked both reliable and valid to us. We didn't

have concerns there.

CHAIR SEPTIMUS: Any other comments, since we had partially discussed this. But is there any other comments from the committee? Peter.

MEMBER HAVENS: I want to thank the developers for putting in a kappa statistic under the results section where it is easy to find and notice that it is 0.47, which suggests moderate reliability.

DR. WONG: Thank you.

CHAIR SEPTIMUS: Any other? Aaron.

MEMBER MILSTONE: I just had a question about the reliability of the CPT category II codes across different systems and types of providers because this is looking for where they were receiving therapy, it was did someone actually document that they were giving therapy but not looking for the drug itself?

DR. WONG: I am going to punt.

MS. CHRISTENSEN: Can you ask the

question for me one more time?

MEMBER MILSTONE: So we discussed this in a couple of measure in our workgroup where we saw the CPT category II codes as a measure -- as a marker for whether or not a patient had either gotten a test or a drug. And here you have in your denominator the CPT Category II Code for patient receiving antiviral treatment for hepatitis C. So it doesn't mean that the patient is on a drug. You are not looking for a drug in the med list. You are looking for did a provider or coder check a box that led to that code being --

DR. WONG: It is the reliability of CPT II Codes across health -- across EHRs.

MEMBER MILSTONE: So I'm trying to figure out like how reliable that CPT II code for patients on antiretroviral therapy.

MS. CHRISTENSEN: So our testing project is pulling information from an electronic health record, which that would not be CPT II Codes. That would just be clinical

data indicating that something was done or not done. Does that make sense?

DR. WONG: So there would be a procedure code that you billed for that was a viral load as opposed to a CPT II Code.

MS. RALLINS: Yes and in addition to that, we would also use the RxNorm. So there are clinical vocabulary codes that we use to capture that in an electronic health record.

MS. BURSTIN: But I think that -- I'm sorry, just to interject. I think what is being asked is what is the testing and reliability of the measure based on CPT II Codes and that is not available. At this point, we only have testing based on the EHR.

MS. CHRISTENSEN: Yes, so we do have one study of the measures being tested and used in claims but it was not a project designed to test the reliability of the measure. It is more testing performance across different patient groups with disparities. So we didn't provide that information because I don't think it is

really relevant to the question that you are asking. But they did do testing. It just doesn't break it out the way you would want to see it for this question.

MEMBER MILSTONE: So let me try to clarify my question. So the numerator as I read it is within six months prior to the initiation of antiretroviral therapy. All right, you have to be a new initiate and have a hep C. So your denominator or people that should include all people that are newly initiated on antiretroviral therapy. Those are eligible and then for having viral load testing.

But then when I looked down to the denominator details, it doesn't have a -- it doesn't say how you are capturing patients who were newly -- I'm sorry. I'm looking at 2a1.7, right there.

So the denominator details, unless it is somewhere else, how are you identifying people that are newly diagnosed with hep C?

MS. CHRISTENSEN: Okay.

MEMBER MILSTONE: I'm sorry. Sorry -- who are newly initiated on therapy.

This gets back to questions we asked earlier. How do you know you are measuring what you want in the population you want to measure it versus you are only capturing -- right now you are only capturing this in people that have this CPT Category II Code for patients receiving antiretroviral therapy.

Maybe I am missing something. Please, chime in.

MS. CHRISTENSEN: So if you are looking at 2a1.7, the EHR specifications are attached and then the claims specifications are listed there for you, if that helps clarify it. So that those are the EHR specifications that we are looking at now. And if you scroll down a bit,, there should be, I think, a list of data elements and then maybe a logic diagram.

DR. WONG: If I could just say that all of these measures that we are proposing have been in use and people have used them. So it

is definitely feasible.

MEMBER MILSTONE: But this is again about reliability and validity so it is how confident are you that you are identifying the population that you think you are identifying using this measure both in EHR and not in EHR. Because this applies for EHR but it doesn't for people that are still on --

So if you have previous data, because this is not a new measure, you said you don't know how well it works outside of EHR.

So again, how well can it identify patients newly initiated on therapy that aren't covered under an electronic health record?

MS. RALLINS: Okay, so your questions is -- because I am looking at the specifications for EHRs. And your questions pertain to claims or to EHRs?

MEMBER MILSTONE: I guess primarily claims.

MS. RALLINS: Okay, so for claims we would use the CPT II Codes. But I can't say

that the CPT II Codes identify the newly diagnosed patients because the CPT II Codes are not written that way. That is what you are asking, right?

MEMBER MILSTONE: So you are saying that the CPT Codes haven't been validated to detect people in the denominator.

MS. RALLINS: So I don't have the CPT Codes in front of me. So you want to know --

MEMBER MILSTONE: I guess I'm not asking a clear question.

MS. RALLINS: So you want to know how the CPT II Codes identify the newly diagnosed patients or patients that are newly --

MEMBER MILSTONE: Newly initiated on treatment.

MS. RALLINS: -- initiated on treatment.

MEMBER MILSTONE: So if you don't have an EMR.

MS. RALLINS: Right, I get it. So

what I am saying to you is we need to look at those CPT Codes. Can we come back to your question?

MEMBER MILSTONE: Sure.

MS. RALLINS: Yes, we can do that.

DR. WONG: I don't think that is the question. Perhaps I am wrong. But you are asking not newly diagnosed with hepatitis C. You are asking --

MEMBER MILSTONE: Newly initiated treatment.

DR. WONG: -- newly treated on therapy.

MEMBER MILSTONE: Which is the denominator that you have in your measure. Newly initiated on therapy.

MS. BOSSLEY: Right. So, Keri, you haven't yet tested this measure using CPT II, correct?

MS. CHRISTENSEN: Correct.

MS. BOSSLEY: So you honestly can't answer whether that CPT II Code that is received

prior to initiation of therapy is indeed reliable, when it is --

MS. CHRISTENSEN: Well we need to look at the language in the CPT II Code before we answer it.

MEMBER MILSTONE: Can you go up on the screen? Because there are actually two parts to the question. So let me try to clarify again. I'm sorry. Can you go up a tiny bit.

Because there are two issues. I mean if you have someone who is on EHR, if you have a system, if you validated this within an electronic health record system, the question there is how well is this -- what is the reliability of this at detecting those patients, the denominator patients using this algorithm. And the second is, if you don't have an EHR, how well can you identify those that were initiated on antiretroviral therapy using those CPT Codes.

MS. CHRISTENSEN: Right. So the answer is to the first question, that should

be reflected in the reliability of the measure that we gave you because we would have checked that as part of the abstraction to ensure that they met the qualifications for the measure.

So the reliability of the denominator would be the same as or higher than the reliability that we provided.

And then the second question, the CPT II Code should only be used to indicate when it is one of the patients that meets the measure. But as Heidi correctly pointed out, we did not specifically go back and validate claims, which is very difficult to do because, frankly, providers don't really like you to go back and validate their claims. Again, it is the CMS fraud and abuse part.

So we have tried in some projects. We did not do that in this project.

CHAIR SEPTIMUS: That is probably as clear as mud to everybody.

MS. RALLINS: So when looking at the CPT II Code that is there, it isn't clear if

you can identify the patients that are receiving newly, patients that have just been placed on the drug. Is that what you are asking?

So the CPT II Code is patient receiving antiretroviral treatment for hepatitis C. So it is only used for this measure. You know, it doesn't -- we haven't tested it but I would presume that regardless of if you have been on the drug for a while or you have just been placed on the drug, this code could be used for that but we haven't tested that yet.

CHAIR SEPTIMUS: Helen, do you want to comment?

MS. BURSTIN: Just a general point. We only endorse measures on the data platforms on which they have been tested. So essentially you have only been provided testing data at this point on the EHR specs with reliability so I think your questions are very valid. And I think we would not be, at least at this current time, endorsing the CPT II-based specs because

we don't have testing on them.

MEMBER MILSTONE: So is that something that you -- I mean, can you say that you endorse it based on EHR and not --

MS. BURSTIN: Yes, the e-specs. Correct.

MEMBER MILSTONE: Would that come out of this?

MS. BURSTIN: Yes.

MEMBER MILSTONE: Should CPT not be in here?

MS. BURSTIN: Yes, correct.

CHAIR SEPTIMUS: Is CPT, I mean not CPT, but is the Category II codes there for showing the gap or is it there as part of the measure? Okay, well then that is -- then we have to -- then correct me if I am wrong. We have to vote on what has been presented to us.

MS. BOSSLEY: So what we can do is ask PCPI to make a modification to the form that they remove any specifications related to the claim CPT II and that it only remain specified

for an EHR because that is the testing that you have before you. And then if that is agreed to, which I think they are, your voting should reflect what you are presented related to EHR.

CHAIR SEPTIMUS: Aaron, are you finished? I guess we have to go to Tom now.

MEMBER FILE: Whatever. I just want to be clear on this so that we are politically correct or whatever. But do we have the ability to change their process that they are presenting to us? I mean it seems to me that that is not our responsibility.

MS. WINKLER: Yes you do because what you have been presented are two versions of the measure, one of which is tested and one of which is not and you can say we don't know enough about the one that is not tested to say anything. We can make conclusions and recommendations based on the part that has been tested.

MEMBER FILE: All right, so this has to be amended somehow for our vote.

MS. WINKLER: Yes. We'll take care of that, yes.

CHAIR SEPTIMUS: All right. All right, so what we are going to be voting on now about reliability is the e-spec and that is a Category II, correct?

MS. WINKLER: Correct.

CHAIR SEPTIMUS: Okay, so are we ready to vote then on that revised spec? Okay then, let's vote.

MS. KAHN: Voting on 2a, reliability; high, moderate, low, or insufficient. You can go ahead and start.

(Pause.)

MS. KAHN: We have one high, 17 moderate, one low, and one insufficient evidence.

CHAIR SEPTIMUS: Okay, the next is validity.

MEMBER CAMPOS-OUTCALT: Again, we did not have as a group any concerns about that. I think the kappa statistics have already been

presented.

CHAIR SEPTIMUS: Okay, keeping in mind the revision and what we are voting for, are we ready to vote on validity then?

Okay, let's vote on validity.

MS. KAHN: Voting on 2b, validity; high, moderate, low, or insufficient evidence. Go ahead and start.

(Pause.)

MS. KAHN: We have zero high, 19 moderate, zero low, and one insufficient evidence.

CHAIR SEPTIMUS: Okay, next is usability.

MEMBER CAMPOS-OUTCALT: This measure has been in use already for it looks like four years and we were presented with nothing that made us question its usability or feasibility for that matter. So both of these we didn't have any concerns about.

CHAIR SEPTIMUS: Any problems with the previous four years, John?

DR. WONG: No.

CHAIR SEPTIMUS: Any comments from the group? It sounds like we can do both -- well, we can't do them together but we will do usability and then as soon as that is finished we will do feasibility unless anybody else has any comments. We will do two right in a row. How about that?

Okay, we will start with usability.

MS. KAHN: Voting on usability, again, high, moderate, low, or insufficient evidence.

(Pause.)

MS. KAHN: So we have 12 high, eight moderate, zero low, and zero insufficient.

MEMBER CAMPOS-OUTCALT: Okay, if you will put up the voting now for feasibility.

MS. KAHN: Voting on feasibility. Again, it is high, moderate, low, or insufficient.

(Pause.)

MS. KAHN: I think we need one more

person.

(Pause.)

MS. KAHN: So we have nine high, 11 moderate, zero low, and zero insufficient.

MEMBER CAMPOS-OUTCALT: Okay, so then the last vote on this measure is suitability for endorsement. Are there any comments on that? Then we will vote.

MS. KAHN: Voting on overall suitability for endorsement, does the measure beat NQF criteria for endorsement yes or no? You can go ahead and start.

(Pause.)

MS. KAHN: Can we have everyone enter their vote one more time? We have 19 yes and one no.

CHAIR SEPTIMUS: The measure passes. So we will now go to 0396, which is related to genotype.

MEMBER CAMPOS-OUTCALT: Okay, so this measure is percentage of patients aged 18 years and older with a diagnosis of chronic

hepatitis C who are receiving antiviral treatment for whom HCV genotype testing was performed prior to initiation of antiviral treatment.

You know, this basically all the way through is pretty much the same as the last one. I mean, I don't know that this is going to require a lot more discussion. I don't recall anything here that was different from the last one, even the kappa statistics on the validity were pretty much the same. It is pretty much the same as the last measure.

CHAIR SEPTIMUS: Any issues around type II coding with the genotypes?

MS. WINKLER: In general, I think we have to look at the whole suite of measures from PCPI in terms of their testing for hep C for their testing having been done only in EHRs and really since that is really all we know about, I think we have to apply the same conclusions that you have already applied to the last one to all of the measures, even the

ones that have gone before. Does anyone disagree with that?

CHAIR SEPTIMUS: That was the reason for the question. Okay, so let's go to -- let's still go through the same thing. Let's go to impact. Is there any discussion on impact then? If not, we will vote.

MS. KAHN: Voting 1a, high impact. Again high, moderate, low or insufficient. You can go ahead and start.

(Pause.)

MS. KAHN: Could we have everyone press it one more time?

(Pause.)

MS. KAHN: We have 15 high, five moderate, zero low, and zero insufficient.

CHAIR SEPTIMUS: Okay, the next is going to be the evidence. Any further comment, Doug, on that?

MEMBER CAMPOS-OUTCALT: For each of these if you just add we have nothing as a group that we came up with.

CHAIR SEPTIMUS: Comments from -- Raymond.

MEMBER CHUNG: Okay, I have a question. This was a genotype obtained any time prior to initiation of antiviral therapy. Correct?

DR. WONG: Yes.

MEMBER CHUNG: Okay. So you are searching the entire database. Okay.

CHAIR SEPTIMUS: Additional comments? Okay, we will vote on the evidence.

MS. KAHN: Voting on 1c, yes, the body of evidence meets the guidance; no, the evidence does not meet the guidance; and no, insufficient information was submitted. You can go ahead and start voting.

(Pause.)

MS. KAHN: We have 15 yes, the body of evidence meets the guidance, four no, the evidence does not meet the guidance; and one there was insufficient information.

CHAIR SEPTIMUS: Well on that last

vote we found that it was Aaron.

(Laughter.)

CHAIR SEPTIMUS: Okay, opportunity. Any discussion on opportunity? Okay, we will vote.

MS. KAHN: Voting on 1b, performance --

CHAIR SEPTIMUS: Raymond, did you have a question? I'm sorry.

MEMBER CHUNG: -- it look like about 86 percent.

DR. WONG: Yes, about 80 -- well, it was 79 percent. But still it is still a gap and there is a huge difference between you are going to treat genotype 1, 2 or 3, as you well know.

CHAIR SEPTIMUS: Kathleen?

MEMBER BRADY: Is there updated data for this measure as there was for the last one for 2010?

MS. CHRISTENSEN: Very similar to what you have got up there.

MEMBER HAVENS: So similar in that you saw the dramatic drop that you saw with the overall viral load or similar to --

MS. CHRISTENSEN: It is similar to the numbers that you are seeing on the screen. No change.

MEMBER HAVENS: Arguing again that the viral load data that you presented prior is a measurement error that has nothing to do with what people are actually doing because you don't get a genotype without getting a viral load.

MEMBER CHUNG: Well that's not necessarily true because this is -- the genotype at any point in that patient's history.

MEMBER HAVENS: Okay, thank you.

CHAIR SEPTIMUS: Additional comments on performance gap and opportunity?

Hearing none, we will vote.

MS. KAHN: Voting on 1b, performance gap. It is high, moderate, low, or --

CHAIR SEPTIMUS: Don't vote yet.

MS. KAHN: -- insufficient evidence.

CHAIR SEPTIMUS: Now you can vote.

MS. KAHN: All right, when the clock starts, you can start pressing the button.

(Pause.)

MS. KAHN: Okay, you have three high, 16 moderate, one low, and zero insufficient evidence.

CHAIR SEPTIMUS: Okay, moving right along, we will go to reliability and validity. We will start off with reliability.

Any comments?

MEMBER CAMPOS-OUTCALT: Again, we didn't find anything new here compared to the last measure.

CHAIR SEPTIMUS: Any comments from the committee? Okay, well, I guess --

MEMBER HAVENS: Yes, thank you. Is there a reason that the reliability results are put in the validity section?

MS. BOSSLEY: So this is part of what our testing task force looked at a couple years ago and generally speaking, the thinking was that if you are testing in an EHR, the reliability, the repeatability is not really what you are looking for. You are looking at the validity. So you are looking at what is produced in the report out of the EHR and back into making sure it can be identified in the EHR. So that is validity. That is not reliability as it is defined by our criteria. So that is why you see it provided in the validity section. Does that make sense?

MEMBER HAVENS: Well, as described, potentially but then there is no validity measure possible?

MS. BOSSLEY: I'm sorry, I didn't catch that last part.

MEMBER HAVENS: Well you suggested that the way you do the measurement in an EHR is not really a measure of validity. That is why you put that answer in the other section.

Is this valid? And where do I go to find the measurement of validity in this document?

That should be in 2a2.3, reliability statistics. It says go down below. Your answer suggested to me there is no way to measure the reliability in an EHR review. That is what you just said, unless I misunderstood.

MS. BURSTIN: No. What we are saying is that the CPT II Codes weren't tested. But what they have done for reliability of the EHR is they do a computation of the EHR, given the structured elements and they do a visual inspection of the record. That is the reliability to see if it wasn't in a structured field and may have been in free text to get a sense of the reliability of the EHR base specs.

MEMBER HAVENS: Then why isn't it reported in reliability? Why is she calling that validity? I am just trying to understand. It is just the way we do it?

MS. BURSTIN: It sounds like it's our fault.

MS. CHRISTENSEN: It is the way NQF has asked for us to present the information.

MEMBER HAVENS: Right. So now I'm asking NQF. I am just trying to understand. That's all.

CHAIR SEPTIMUS: Let me ask this, Peter, seriously. I mean I understand your confusion. Is there -- based on that, do you have a concern about the reliability or validity of this particular measure?

MEMBER HAVENS: No. I'm trying to understand. If we are asked to make criterion-based decisions about these things, and you go to section 2a2.3 where it supposed to give you the testing results for reliability and it refers you down to another section on validity and the answer I get is you can't -- when you do this measurement in an EHR, you can't really measure reliability. It is only a measure of validity. I don't understand what you are saying. That's all.

MS. BOSSLEY: I understand. Helen

and I are not speaking the same language today and I am sorry about that.

So I just pulled out some information. Let me read it because perhaps I am not describing it well.

So when the testing task force looked at this, reliability is looking at the repeatability of getting the same data elements and same score --

MEMBER HAVENS: Right. So when you --

MS. BOSSLEY: -- which they felt was not really needed for an EHR. Because once you code it into an EHR, this was the thinking of the task force, so they then said look at validity. And validity analyzes agreement between data elements and scores obtained with data exported electronically using the specifications to those obtained by review and abstraction of the entire EHR. So you are looking at what is produced out of an EHR in a score of reports and you go back in and look

to see are the results valid. What you produced out of that EHR are valid for what is documented in the EHR.

MEMBER HAVENS: That is --

MS. BOSSLEY: So they defined that as validity.

MEMBER HAVENS: That is a non-standard definition of validity. The reliability measure, which has been shown, for example, in reviews of the VA record where you look at some standard way of extracting something versus smart text extraction shows that you can get more reliable definitions using smart text extraction that you can in other ways. So that would say reliability, but depending on how you extract it from an EHR, is 95 percent. I would take that as reliability. The question of validity is is what you are extracting in these two ways that give you a similar answer really showing what you want it to show. And you have the face validity, which I understand. You can't maybe take that to a deeper level

other than face validity. It's okay.

MS. BOSSLEY: And there is disagreement across groups and that is why you see PCPI actually frames it as reliability.

MEMBER HAVENS: Okay.

MS. BOSSLEY: For the purposes of the discussion today with the criteria we have -- we will send you information -- but as it is defined by our criteria at the moment, it is validity testing. They have satisfied validity testing, which actually trumps the reliability in this instance because it is at the data element level.

So reliability here would be is it precisely specified and then the validity piece would be the testing that is provided with the EHR. And I understand there is a difference of opinion and it is not the first time it has been voiced. But as our criteria stands, this is how they have outlined it.

MEMBER HAVENS: Great. Thank you very much. Then I have no concerns, based on

NQF criteria.

(Laughter.)

MEMBER MILSTONE: I have a quick question about unintended consequences that I will pose to the hepatologists.

So if I am a primary care doctor that is all of a sudden is listening to the CDC and I am going to test all the baby boomers and I am going to identify a lot of patients with hep C and then I am going to say oh, I should get a viral load and a genotype and then I am going to refer them to a hepatologist for treatment, is this going to lead to hepatologists saying well to be in compliance, I am going to have to redo those, so that they are in my medical records. When I get audited, there is a link between my starting treatment and this patient being tested. And this has happened with other things like kappa guidelines and I just want to make sure that this measure linking treatment and testing isn't going to lead to unintended additional testing.

CHAIR SEPTIMUS: That is the attribution issue we discussed earlier.

As I understand it, more than one practitioner will get credit for the measure. So there are some of those if you have gotten that information from another practitioner, you can count that as being done.

MEMBER MILSTONE: So if my EMR doesn't link to your EMR --

MEMBER HAVENS: That would go to the reliability of the extraction measure. It gets to my point.

CHAIR SEPTIMUS: You are absolutely right. We are not tying this to Type II Codes. We are tying this to electronic abstraction.

MEMBER CAMPOS-OUTCALT: Based on that answer, does that make us less sure of this reliability or the validity? Because if it is not -- or I guess -- if we are now thinking that this is going to have unintended consequences, how does that affect our vote?

MS. WINKLER: Unintended

consequences come in under feasibility.

CHAIR SEPTIMUS: And one question I don't know is whether a primary care physician who gets this test is going to do the next two steps that Aaron outlined or are they automatically going to refer this patient to someone who is a hepatologist and I don't have any knowledge of what they are going to do.

MEMBER CHUNG: I think increasingly with time you are going to see more and more of the treatment shift into the landscape above the infectious disease and the primary cares as therapy becomes simpler and non-interferon-based.

So it may actually end up being easier to capture.

CHAIR SEPTIMUS: With protease inhibitors?

MEMBER CHUNG: No, with even simpler agents than that, ultimately.

CHAIR SEPTIMUS: Well maybe down the road but right now, therapy actually has

become more complicated, not less complicated.

MEMBER CHUNG: Yes. Yes, you have to get more complicated before you become less complicated. That's right. That's right.

CHAIR SEPTIMUS: Okay, so the feasibility gets into unintended consequences, which is what I think is what Aaron said. So we will postpone that until we get past the reliability and validity. So but great point, Aaron.

Any other things about the scientific reliability? Yes, Tom.

MEMBER GIORDANO: So given no reliability data, as NQF is instructed, how are we supposed to vote? Is it moderate?

MS. BOSSLEY: Yes, you still need to assess the specifications. Are the specifications provided precise? And that is the e-specification. So to me that would be moderate because high just wouldn't apply in this instance, I don't think. But that would

be really all you are looking at for reliability.

CHAIR SEPTIMUS: Raymond, you -- Okay, let's vote on reliability.

MS. KAHN: Voting on 2a reliability; high, moderate, low, or insufficient evidence. You can go ahead and start.

CHAIR SEPTIMUS: Now we can vote.

(Pause.)

MS. KAHN: You have zero for high, 18 moderate, one low, and one insufficient evidence.

CHAIR SEPTIMUS: Okay, next we will go to validity. Let's see if there is any comments. I think we sort of covered almost both together but just to make sure there is no additional comments before we vote on validity. If not, let's vote.

MS. KAHN: So voting on 2b, validity. Again, high, moderate, low, or insufficient evidence. You can go ahead and start.

(Pause.)

MS. KAHN: So we have one high, 19 moderate, zero low, and zero insufficient evidence.

CHAIR SEPTIMUS: Okay, so now we get into usability. Doug, anything about usability?

MEMBER CAMPOS-OUTCALT: Again, it has been in use for four years and we did not get any information regarding problems.

CHAIR SEPTIMUS: John, any comment about usability from the previous four years' experience?

DR. WONG: Nothing from my standpoint.

CHAIR SEPTIMUS: Anything from the people behind me that I can't see? Comments from the committee?

Okay, let's vote on usability.

MS. KAHN: Voting on usability, high, moderate, low, or insufficient. You can go ahead and start.

We have five high, 14 moderate, zero low, and one insufficient.

CHAIR SEPTIMUS: Okay, now we are going to get into feasibility. And just to remind everyone, this is stuff being generated during care, electronic sources and then number two, the comment earlier by Aaron susceptibility to inaccuracy or unintended consequences identified.

Comments?

(Pause.)

CHAIR SEPTIMUS: I guess we will vote.

MS. KAHN: Voting on feasibility; again, high, moderate, low, and insufficient. You can go ahead and start.

(Pause.)

MS. KAHN: We have one high, 16 moderate, two low, and one insufficient information.

CHAIR SEPTIMUS: Okay, and the last one in this measure is overall suitability for

endorsement. Do we need to have any other discussion? Okay well, let's vote.

MS. KAHN: Does the measure meet NQF criteria for endorsement? Yes or no. You can start your vote.

(Pause.)

MS. KAHN: We have 20 for yes and zero no.

CHAIR SEPTIMUS: Okay, so that passed. Now next Raymond is going to do 0397.

MEMBER CHUNG: This is hepatitis C antiviral treatment prescribed. This is a maintenance of an original approved or endorsed measure from 2008. And it essentially asks for the percentage of patients 18 or older with a diagnosis of chronic hep C who were prescribed at a minimum pegylated interferon and ribavirin therapy within the 12-month reporting period. More on that kind of semantics a little bit later when we talk about minim peginterferon and ribavirin. John alluded to that earlier in his remarks.

From the vantage point of impact, you have heard about the epidemiology, the natural history of hepatitis C. Clearly a number of studies including a couple of dozen studies submitted by the PCPI have demonstrated the salutary effects of a sustained biologic response. That is to say, permanent or sustained clearance of virus on a long-term outcomes and these include particularly liver-disease related outcomes including decompensation, death from liver failure, and hepatocellular carcinoma.

There have been reductions as well in liver-related mortality of magnitudes ranging from 3.3 to 25-fold in one study and a meta-analysis suggesting a decrease in HCC incidents of about two and a half-fold.

So there has been sufficient maturation of data therefore to justify long prevention and postponement of long-term outcomes as the result of obtaining a sustained biologic response.

So the impact of treatment appears to have clear-cut clinical benefits. So with that in mind, this performance measure of documenting therapy in those persons who were deemed eligible and suitable for therapy is really the focus here today. So I guess on an impact basis we could vote on that.

CHAIR BROTMAN: Okay, any discussion on this? All right, Doug, were you going to raise your card?

MEMBER CAMPOS-OUTCALT: Yes. I hate to do it. There was just an evidence report completed on this on intermediate outcomes. Long-term outcomes I don't think have been studied well yet. And the studies are observational. They do all point in the same direction, which is benefit, not of the magnitude that was just mentioned. And long-term I don't think has been studied long enough.

You know, I don't think this is enough to make me vote against treating but I

do think something that was brought up during our phone discussion which bears talking about, which is if I were a patient right now and I knew all these newer and beneficial treatments were coming down the road that might be more benign, would I rush to be treated right now. And I think that is a fair question.

CHAIR BROTMAN: That came up to a very significant degree in our workgroup. Does anyone else have any comments about the new treatments on the horizon and how they would be treated?

MEMBER THOMPSON: I can just tell you like my partner is hep C. A lot of the people in the community doing patient education, we are telling people to wait. That if it is not an immediate need for them, to wait for the new treatments to come out. So I think that is a completely legitimate concern.

CHAIR BROTMAN: Tiffany?

MEMBER OSBORN: How imminent are the new treatments that are coming out?

MEMBER CHUNG: We are looking at perhaps the first -- well, we already have new treatments in the form of add-ons, in the form of telaprevir and boceprevir but they are piggybacked on to peg and ribavirin.

When we are talking about all oral combinations, which is really what Adam is getting at, we may be looking at the first combination, at least for genotype 2/3 infection in about a year and a half. Phase three is just about completed. Enrollment is completed, I should say of the Phase III study of the two-drug oral combination for genotype 2/3 infection. Soon to follow will be deep phase studies for genotype 1.

So the short answer may be anywhere from one and a half to the next three years or four years for roll out of all orals for all genotypes, presumptively. So it is a short time horizon.

CHAIR BROTMAN: Please.

MEMBER GIORDANO: Yes, we did

discuss this in detail and my recollection was the biggest concern I had and other people had was that it wasn't clearly delineated in the denominator that this could be an exclusion that a provider or patient decision could be to defer the therapy. And my understanding from what I think John responded to it is that would be included. And I just don't know, everybody else seemed to be hung up on the same thing, if the language of that might be able to be modified to make sure that that exclusion is validated in the measure.

DR. WONG: So in the measure, as stated we would consider it a medical exclusion, so that the patient perhaps because of a low fibrosis level and also in discussion with their provider could opt to postpone treatment until the newer agents become available.

In our concept of medical exclusion, that would be an appropriate treatment for the patient for medical reasons because the stage of fibrosis was low, or the patient opts to wait

for an all-oral agent combination of drugs, or for other reasons.

I'm certain we could, if the group felt that it was appropriate, make that more explicit. I can understand why some physicians might not consider that a medical exclusion. In general terms, we would, and we could, I'm sure, reword it to say for example. Again, we wanted to be very careful about not over-specifying the reason for medical exclusion. But since this is an issue that some physicians may misunderstand, I think we certainly can clarify the language.

CHAIR BROTMAN: Raymond, go ahead.

MEMBER CHUNG: I think from a disease -- you know, I think you could call that a disease management exception of some sort or exclusion. I think this fundamentally gets at a divide that perhaps separates most of the docs in this room from what I am, which is to say that fundamentally hepatitis C, until now, has been a liver disease. It has been a liver

disease because the therapies have been unpalatable and so we treat liver disease because it demands it. It demands it clinically. There is sufficient advancement of disease.

But with the lowering of the threshold for therapy, we are increasingly moving to the arena of making it an infectious disease. It is now treat the virus for the virus' sake, irrespective of disease, stage because your threshold is low. Your barrier to treatment is decidedly diminished.

And so I think that is what we are -- it is a very moving target. And in four years, this paradigm will be appropriate when you just say instead of peg ribavirin, you just say approved antiviral therapy. And maybe that is what you should be saying even now. And honestly, I just, I think that there should be a disease management exclusion at least until such time because I think we are still in a peginterferon world right now. And that is a

justifiable reason not to treat. Otherwise, you are going to have physicians fall short on these performance measures year after year after year because they have chosen not to treat.

CHAIR SEPTIMUS: It is nice, Raymond, that you now know that really everything relates to infectious diseases.

(Laughter.)

MEMBER CHUNG: I've been coming around to it.

DR. WONG: Just very briefly, I agree that the issue is for us to propose a quality measure we would have to have a recommendation along those lines. And as such right now the recommendations across the various guidelines are at most or at best triple therapy. So until those new agents are approved, in addition, as will come up, there will be I think an issue of specifying what is acceptable antiviral therapy, as has occurred within the anti-HIV treatment regimens. But again, we agree that this is a placeholder until the new

agents emerge and it will be a challenge going forward to specify those.

CHAIR BROTMAN: Tiffany?

MEMBER OSBORN: I'm sorry, I'm still just trying to understand. And it could be because it is the end of the day and it was a very stressful morning, wasn't it?

So what I am trying to get is that we are saying that we have the specific therapy to treat hepatitis C but that therapy is going to change within the next three years and this will come out in 2013.

So I'm trying to understand why we would say we would penalize people for not --

(Laughter.)

CHAIR BROTMAN: We have some comic relief going on in here. Excuse us.

MEMBER OSBORN: I am trying to understand why we would penalize physicians for either waiting or electing to use the oral therapy rather than -- why do we have this? If there is another therapy coming out within

three years that the patients clearly seem to prefer, why would we penalize physicians for not using that?

DR. WONG: So we would not necessarily be penalizing them. That is the rule for the exclusion. So if the physician documents either any one of these three: a medical reason not to give therapy now, which could include that you see a bunch of new drugs on the horizon and you don't have very bad liver disease; or it could be a patient preference, meaning that the patient doesn't really want treatment now. That would also be an exclusion. So again, the doc is not penalized for any of that. Or it could be a system level exclusion, meaning that the insurance company requires a high copay for these very expensive drugs.

So again, we try to be very careful to allow physicians when it is inappropriate not to necessarily administer antiviral therapy. The reason to have this is related to hepatitis C morbidity and mortality.

Now clearly, the RCTs demonstrate increased SVR with currently available antiviral therapy. Secondly, if you just look at the folks who have SVR, and the reason the Backus study is so often cited is because it is the very best study we have as of 2012 that links an indirect outcome or a surrogate outcome, which is sustained viral response, to a whole bunch of long-term outcomes. Meaning that you remain viral negative and you have reduced all-cause mortality and hepatocellular carcinoma and decompensation.

Now, I admit that none of the RCTs that were used for registration trials have gone out and measured and followed their patients over ten to 20 years, which is what is required to have hard outcomes from that. However, the interferon study that I provided you involves RCTs from patients who just got interferon 20 or 30 years ago. And when you look at interferon versus no interferon, there was a statistically significant reduction in hepatocellular

carcinoma.

MEMBER OSBORN: So am I understand correctly then really the crux of what you are trying to accomplish is treat with one of these different options, unless the patient doesn't want to be treated or -- so maybe instead of just saying that it has to be ribophorin or interferon and interferon that they have more choices than that?

DR. WONG: Right. So for genotype 1, the recommended therapy is pegylated interferon, ribavirin, plus one of the new industry for a protease inhibitors, either telaprevir or boceprevir.

Now we could have split this up into multiple measures, meaning that for genotype 1 you get X, for genotype anything else you get Y. That would require an EHR and extraction of the particular genotype, a painful process documenting it, observing it. We see that there are all these new treatments down the pike. This is, as I said, I anticipate this to be a

placeholder. We will be back with new proposals probably within the next year to two. My guess would be two or three because they do have to be reflected in guidelines. But again, for those patients with advanced fibrosis, there are those who would benefit from a demonstrated efficacious therapy and this measure is to try to encourage that.

CHAIR BROTMAN: John, I'm sorry to cut you off. We have to get back on track. Just three quick comments, if you have anything very quick.

MEMBER CHUNG: Very quickly. To Tony's point, which is that right now the pie is loaded with exclusionary slices and you only have 20 percent of that pie actually being treated and you are evaluating performance in that 20 percent slice. It is kind of almost in a recommendation before its time at some level, based on these arguments. I wonder if it makes sense for this to be, rather than documentation of treatment, a documentation or

performance of having had a discussion about antiviral therapy with every one of these patients and a disposition therein that reflects the time, the moment, the period.

DR. WONG: I think it is a great proposal. The issue is that we have sort of -- and this will come up later when we talk about counseling, is documenting those measures, the potential for gaming. If you are truly interested in outcomes, which is what we were, this is as close as we get.

CHAIR BROTMAN: Okay. Adam, do you have something quick to add?

MEMBER THOMPSON: I just wanted to add that when you are looking at the denominator exclusions, I am really uncomfortable with the third one. I think it is a really nice way of saying that my patient was poor, so I don't have to be held accountable for not prescribing them medication by saying they don't have insurance or the therapy is not covered and that they get pulled out of the denominator. I mean, I think

that speaks to systems having to be held accountable for accessing treatment for their patients. So I just wanted to throw that out there.

CHAIR BROTMAN: Okay. We are going to have probably more discussion after this but I think we ought to vote on the impact portion at this point and see where it goes.

MS. CHRISTENSEN: Can I just respond to that?

CHAIR BROTMAN: Sure, go ahead.

MS. CHRISTENSEN: I just want to remind everybody that that data is not lost. The patients who are exceptions should be tracked and reported alongside. And in the PQRS program they do track and report those not publicly but they do track and report them so that you are not losing the data. It is seen, the exception data.

CHAIR BROTMAN: Okay, so let's go to a vote on impact.

MS. KAHN: Voting on high impact;

high, moderate, low, or insufficient evidence. You can go ahead and start.

(Pause.)

MS. KAHN: We have ten high, five moderate, four low, and one insufficient evidence.

CHAIR BROTMAN: Okay, so that passes. Yes, let's go to evidence at this point.

MEMBER CHUNG: I think that I have covered a number of the studies that actually have been presented in two different formats. One, the evidence provided from the PCPI supplement, as well as the documents relating to physician statements from a variety of organizations, ASOB, practice guidelines, ACP, and a number of other associations, with level 1a evidence to support not only the superiority of antiviral response rates or of currently approved therapies over preexisting treatments but also clearly speaking to the clinical benefits that I alluded to earlier.

So the evidence to support the benefits of treatment, I think, are strong.

CHAIR BROTMAN: Any discussion on the evidence presented? Sure, Tiffany.

MEMBER OSBORN: Just one question and that is, given the fact that you have talked about the importance based on the severity of illness, how does the fact that there is no risk adjustment or risk stratification impact this?

MEMBER CHUNG: I'm sorry risk stratification for?

MEMBER OSBORN: So previously you discussed the fact that especially if they --

MEMBER CHUNG: So first stage liver disease?

MEMBER OSBORN: Yes.

MEMBER CHUNG: So most of these trials were actually conducted in patients who had a mix of liver disease ranging from early fibrosis and generally speaking randomized controlled trials usually a segment of say anywhere from 10 percent to perhaps more

cirrhotics.

So cirrhotics were usually included in the randomized registrational trials of the compounds we have spoken about earlier. There have been some cirrhosis and bridging fibrosis directed trials but for the most part, they have been incorporated as minority components. And the response rates have also been, while somewhat attenuated, have still been excellent in those groups in terms of those persons who were naive to treatment.

CHAIR BROTMAN: Any other discussion? All right, let's move to a vote on the evidence, please.

MS. KAHN: Voting on 1c, evidence, yes, the body of evidence meets the guidance; no, evidence does not meet the guidance; or no, insufficient information was submitted. You may begin your vote.

(Pause.)

MS. KAHN: We have 13 for yes, the body evidence meets the guidance; six for no,

it does not meet the guidance; and one for no, insufficient information was submitted.

CHAIR BROTMAN: Okay, so passes. Let's go back to performance gap.

MEMBER CHUNG: Performance gap vantage point because this has been a PQRS program since '08, there are data available and that gap would appear to be about 68 percent performance rate. So clearly, there is room to move.

In terms of again, this is the eligible treatment population, once you winnow out all those exclusions, at least my interpretation. So of those eligibles who don't have contraindications, who haven't opted out, who haven't had physicians decide this is not the time for them, 68 percent met the performance measure.

MS. WINKLER: Just a question. Do we have any information on disparities, since we know this is a big issue?

MEMBER CHUNG: There are let's see,

and John you may be able to clarify this more, but there were some data certainly on disparities in terms of the prevalence of hepatitis C among African Americans, particularly double the rate seen in Caucasians. There is also really sort of the double whammy evidence of a halving of the response rate in African Americans with peginterferon ribavirin therapy. That gap is narrowed with the addition of telaprevir boceprevir in genotype 1 patients. But still there is a gap in terms of success of therapy.

So you are looking at proportionally more minorities infected with lower performance rates in terms of antiviral therapy. So there is a real need.

CHAIR BROTMAN: Kathleen.

MEMBER BRADY: To follow up on that but not just how the drugs perform in ethnic minorities but is there information regarding receipt of therapy by racial ethnic minorities?

MEMBER CHUNG: Yes. I can't cite

chapter and verse the paper but I can tell you that intercity populations disproportionately loaded with ethnic minorities have exceptionally low treatment rates with peginterferon ribavirin. So studies from urban hospitals, if you look at this sort of tree of 100 patients entering a clinic, at the end of the day, less than five percent of those obtained in real world terms a sustained biologic response. I think it was two percent at the end of the day. And when you winnowed out all the exclusions, the preexisting conditions, the contraindications, the lack of social supports.

CHAIR BROTMAN: Tiffany.

MEMBER OSBORN: You may have just said this but does that also count for access to care?

MEMBER CHUNG: Yes, that would be very much, I think, structured into some of the analysis that I referred to in some of those urban studies.

CHAIR BROTMAN: Okay, if there is

no other discussion, let's vote on performance gap.

MS. KAHN: Voting on 1b, performance gap; high, moderate, low, or insufficient evidence. Go ahead and start.

(Pause.)

MS. KAHN: We have seven high, 12 moderate, one low, and zero insufficient evidence.

CHAIR BROTMAN: Okay, so that passes. Reliability.

MEMBER CHUNG: From the reliability vantage point, in terms of structuring our numerators and denominators, we have already had, I think, a bit of a discussion about what actually constitutes that measure. The numerators, I think are straightforward enough in terms of prescription data but the denominator is what, again, demands I think clarity and granularity in terms of our description. And perhaps a little more, instead of this medically excluded -- I

understood there is systems exclusions, there are patient exclusions, but this medical exclusions I think should be perhaps stratified a bit further to medical contraindications or relevant contraindications but maybe also talking about disease management decisions and differing therapy. And so I think that that would be one of the elements of concern about the denominator structure.

CHAIR BROTMAN: Peter?

MEMBER HAVENS: Can I ask a question? Who did the testing on that reliability or validity? It is kind of lower than I would have expected to see. When you look at an EHR generated statement versus a chart review of the same electronic record, I assume that is what you are doing. Right? And the kappa was not as high as I might have expected it to be, based on what you would think would be the same data.

So do you have a reason for that?

MS. CHRISTENSEN: Yes. So we

actually asked them to go in and do the testing on their system as it stands. And then they go back and they make work flow changes because the groups we work with do actually continue to use these measures after they test with us. They don't just do testing for testing's sake.

So if we went back today, likely reliability would be higher because of changes they have made to their EHR just to better capture data.

MEMBER HAVENS: Right. So when NQF approves this, do they approve it for the first pass reliability or validity or for the reliability and validity that might be interpolated based on what you have just said that after somebody goes back and checks you, you actually get better because you change your EHR to really capture the data.

This is important. As a physician who is getting measured and not paid, what you just said shivers my bones.

MS. CHRISTENSEN: All right. So

another thing to point out. It is a great, great point. We would love to go back and test on a regular basis. It is very expensive, you know, $50,000 on this. But a very, very important point -- it pays my salary.

It is a very, very important point that electronic health record automated reporting consistently under-reports performance, unless you go in and make the changes to your EHR to be able to capture data in a way that you can report it out. So that is an important point. There is a motivator there to capture date better to be able to do better on the reporting.

MEMBER HAVENS: So as prescribed in this current document, this would under-report, as proven by your studies, it would under-report the adequacy of physician practice by an unspecified unmeasured amount.

MS. CHRISTENSEN: Unless you go in and make changes to your system. I mean, that is all measurement. If you measure poorly --

MEMBER HAVENS: No. No. Excuse me. Wait, wait, wait. Wait, wait, wait, wait wait. That's like saying it is all words. Right.

We are talking about measurements. And the measurements that you are suggesting that people do lead to a kappa of about moderate at best. If you study it, you have said that the kappa may double -- you haven't said. You said it increases substantially. You said as written it under-reports physician practice. And so I am asking --

MS. CHRISTENSEN: But that is the literature not just our measures, not just this measure, not just these hepatitis C measures. The literature as a whole shows that.

MEMBER HAVENS: So then should we approve this only based on the second pass through an EHR, if that is the only accurate way? No, we shouldn't. We should approve it as written. Go ahead. You have the floor.

MS. CHRISTENSEN: It would

theoretically be possible to design a system to have 100 percent --

MEMBER HAVENS: Well, I'm not talking about 100 percent.

MS. CHRISTENSEN: We are trying to show the real state of the world by going into a site that hasn't gamed the system, if you will, to get a perfect score.

MEMBER HAVENS: I'm not talking about a perfect score. I was impressed at the relatively low kappa statistic on something that seems like it is mom and apple pie.

And you are telling me that after you look at the EHR, you can actually bring up that kappa statistic, suggesting that the way that this is --

MS. CHRISTENSEN: By asking providers to start documenting information or document information in a different way, yes.

CHAIR BROTMAN: I think it is a significant discussion but I mean we have to view the document and the submission the way

it is.

MEMBER MILSTONE: So we are saying this is an e-Measure. So I have a question about the denominator exceptions and I was wondering how many electronic medical records contain these field codes.

So in the data set you validated, they must have these field codes for exceptions to why patients haven't got or shouldn't be on therapy. But I am wondering how many -- like in my or someone else's EHR, there is field codes available and can we even detect them or exclude them from the denominator?

MS. CHRISTENSEN: That is a great question. We meet with a variety of different EHR vendors from the Electronic Health Record Collaborative and discuss these things on a regular basis. So it is an ongoing work with them to help them make sure that they are capturing exceptions correctly. And they provide feedback about how we can develop our specifications so it is easy for them to be able

to capture.

MEMBER MILSTONE: So I guess to follow up on that, do you have a sense of how many companies that provide electronic medical records include these field codes; 10 percent, 50 percent, 90? And I think we all know that to kind of work with translational databases to try to merge data, if the field codes aren't the same, it could be really hard to actually get the data.

So it is a matter of how many have them and are the companies kind of creating these similar field codes that they can really be painted the same way.

MS. CHRISTENSEN: Yes, that is a good question. I don't have any hard numbers to give you but it is important to point out the PQRS program does use these categories. So if they are going to be able to report for PQRS, they would need to be able to capture these categories.

CHAIR BROTMAN: Okay.

MS. CHRISTENSEN: But I don't have a number. Sorry.

CHAIR BROTMAN: Okay, Tom and then we are going to do the other Tom. Tom File.

MEMBER FILE: Okay, very quickly. This is for John. I assume an exclusion would be a patient who has failed prior therapy, let's say maybe two or three, four years ago. That would be a medical -- that would have to be in there.

DR. WONG: That would be a medical --

MEMBER FILE: Is that easy to capture?

DR. WONG: Again, it would be a medical exclusion. It would have to be documented as such within the electronic health record.

MEMBER FILE: All right.

CHAIR BROTMAN: Tom?

MEMBER GIORDANO: So in section 2b3.3, the results for -- I'm sorry. The

CHAIR BROTMAN: Tom. Tom, speak into your microphone.

MEMBER GIORDANO: Oh, I'm sorry. In section 2b3.3, the results for validity, it said the percentage of false negatives due to exception. The number of patients who appeared to fail the measure on automated calculation but were found to not meet the numerator and have a valid exception on the manual review was 46 percent.

So I think as an electronic measure, it is, in my opinion, that is failing validity. And to expect providers to manually document, there is so much expectation that providers are going to document, document, and document, that I don't find that as an acceptable alternative.

CHAIR BROTMAN: So we are still on reliability. That speaks to validity, I believe.

MEMBER GIORDANO: Well isn't electronic measure -- right. So they are intertwined.

MEMBER HAVENS: That's a discussion we had before.

CHAIR BROTMAN: Yes, you got me.

CHAIR SEPTIMUS: You don't want to stir up Peter again, do you?

CHAIR BROTMAN: All right, I think we have had quite a discussion. Let's vote on the reliability and if we need to pick it up for validity, we can pick it up again. But let's vote for reliability at this point.

MS. KAHN: Voting on 2a, reliability; high, moderate, low, or insufficient evidence. You can go ahead and start.

(Pause.)

MS. KAHN: We have zero for high, eight moderate, 11 low, and one insufficient evidence.

CHAIR BROTMAN: So this fails. This is a stop measure. So that is the end.

So we can move on. I think David you have the next one.

MEMBER SPACH: This measure is 0398. This is a maintenance measure. It was initiated originally in July 2008. The measure title is "Hepatitis C: HCV RNA testing at week 12 of treatment." Let me emphasize that the title is at week 12 of treatment. The actual description of it is percentage of patients aged 18 years or older with a diagnosis of chronic hepatitis C who are receiving antiretroviral treatment for whom quantitative HCV RNA testing was performed at no greater than 12 weeks from initiation of antiviral treatment.

So that was one of the issues that was raised, that the title is slightly different than the description; although I think this is a result of this is a measured revised measure and the revised measure is meant to be more inclusive.

The impact of this, we have touched on a number of these issues as we have gone through this but the impact of testing people for treatment results is extremely important

because it will dictate the duration of therapy which has major impact on overall cost of therapy and success of therapy.

So perhaps we can stop there. And John, I don't know if you want to make any comments or we can just vote on it.

DR. WONG: Yes, just briefly. We would be more than willing to rename the measure to within or at 12 weeks.

CHAIR BROTMAN: Any other discussion on it? Okay, let's vote on impact.

MS. KAHN: Voting on 1a, high impact; high, moderate, low, insufficient evidence. Go ahead and start.

(Pause.)

MS. KAHN: So we have 11 high, nine moderate, zero low, and zero insufficient evidence.

CHAIR BROTMAN: Okay, that passes. Let's go to evidence.

MEMBER SPACH: So the evidence is based on a number of studies. Originally the

original document referred to the AASLD guidelines in which this was given class. Because there are different time points of measurements, there were different classes that were looked at; Class 1a, 2ab, and 2b. The documentation was only given in reference to the guidelines but there has been subsequent information that has been provided that was in the word document. And this includes a total of 14 studies in which the antiviral responses in the course of therapy at week 12 or prior to week 12 of therapy had a direct outcome on the subsequent duration of therapy. These studies included at least six meta-analysis and at least four randomized controlled trials, the most notable are three *New England Journal* studies. I think at least two of the three of these were in the New England Journal, which are the SPRINT-2, PROVE 2 and REALIZE trials, which all looked at the issue of response-guided therapy and being able to use virologic responses during the first 12 weeks of therapy

and then dictating the overall course of therapy.

The reason this has such a big impact in terms of overall healthcare is that the cost of hepatitis C therapy is extremely high. The cost and side effects of pegylated interferon an ribavirin are extremely high. And so anything that can shorten duration of therapy can be very important.

And then the other major point with this is that virologic responses at week 12 were being used very heavily now to use so-called stopping rules so people who are genotype 2 and 3 who receive and have viral loads that do not drop more than two logs are considered failures and are stopping therapy. Individuals on telaprevir-based peginterferon ribavirin who don't drop down below 1,000 at 12 weeks are stopped on therapy. People on boceprevir who do not drop below 100 at week 12 are stopped on therapy.

So these particular measurements

early in therapy or particularly at week 12 have a big impact on the overall ability to stop therapy and reduce costs and toxicity to patients.

CHAIR BROTMAN: Any comments regarding the evidence presented?

All right. If not, let's vote on evidence.

MS. KAHN: We are voting on 1c, evidence. Yes, the body of evidence meets the guidance; no, the evidence does not meet the guidance; or no, insufficient information was submitted. You can go ahead and start.

(Pause.)

MS. KAHN: I think we are one short.

(Pause.)

MS. KAHN: Thank you. We have 17 for yes, the body of evidence meets the guidance; two for no, the evidence does not meet the guidance; and zero for no, insufficient information was submitted.

CHAIR BROTMAN: Okay, so that

passes. Let's go into performance gap.

MEMBER SPACH: So performance gap, to my look at this and I will again defer back to John on this because as he has pointed out, I think this is a selected sample. They say the gap and care as shown by this, this is the CMS PQRIS data that says there is a gap in care as shown by this data 89.92 is the aggregate performance rate in the total patient population and 91.63 is the mean performance rate of TIN NPIs. It seems like the gap is relatively small but I will toss that back to John.

DR. WONG: Thanks. In the Kanwal study, it was about 60 percent. And there aren't any data yet available for the triple therapy drugs.

CHAIR BROTMAN: Any discussion on the performance gap issues?

All right, let's go to a vote.

MS. KAHN: Voting on 1b, performance gap; high, moderate, low, or insufficient evidence. You can go ahead and

start.

(Pause.)

MS. KAHN: We have three high, 15 moderate, one low, and zero insufficient.

CHAIR BROTMAN: Okay, so that passes. Let's go to reliability.

MEMBER SPACH: There was significant discussion in the group and on the conference call regarding the reliability and validity, mainly because of the way the measure was worded and we had clarification on this.

The way the measure is worded is that essentially you can get -- it is worded that you need to get a viral load within 12 weeks and the discussion came up and this could maybe generate a little further discussion is that based on genotype 1 patients requiring a viral load response at week four, are looking for a rapid virologic response. It was just a little confusing how precise the measure could be. The response to that was the measure was meant to be inclusive and that is why they chose the

12 week parameter, not to be exclusive. And then John again, you may want to comment on that.

DR. WONG: No, I think you summarized it very well.

MEMBER SPACH: Well Ray, I don't know if you have a comment on that either.

MEMBER CHUNG: No.

MEMBER SPACH: Good? Okay.

CHAIR BROTMAN: Okay, no other discussion. Let's go to vote on reliability.

MS. KAHN: Voting on 2a, reliability; high, moderate, low, or insufficient. We can start.

(Pause.)

MS. KAHN: We have 17 right now. So I am missing one vote.

(Pause.)

MS. KAHN: All right, there we go. So we have zero high, 15 moderate, one low, and one insufficient.

CHAIR BROTMAN: Okay, that passes. Let's go to validity.

MEMBER SPACH: There was less concern about validity. It was felt that the test, the viral load test is a very valid test. Ability to measure that or to be able to extract that from electronic health records is easy to do.

And I don't know if there are any other comments anybody else wants to make on that. We didn't have a lot of --

CHAIR BROTMAN: Aaron.

MEMBER MILSTONE: Just to be consistent with the last one, this one also if you can just expand more on the exceptions. And again, this one brings up the concept of using -- there is EHR specifications but then it also brings up the use of the CPT Category II Codes for documentation of medical reasons for not performing. So I am looking at --

MS. WINKLER: Aaron, we have already determined that for the entire group of measures for hep C --

MEMBER MILSTONE: Oh, okay. So

this is going to -- sorry.

MS. WINKLER: Just the EHR.

MEMBER MILSTONE: That's fine.

CHAIR BROTMAN: Tom?

MEMBER GIORDANO: A quick question for the hepatologist. Would this measure apply, regardless of genotype? Would it apply for genotype 2 and 3?

MEMBER CHUNG: For a peg ribavirin world, if you haven't had a two log reduction in 12 weeks, it is not going to fly.

MEMBER GIORDANO: Yes.

CHAIR BROTMAN: Any other discussion? All right, then let's vote on --

CHAIR SEPTIMUS: I have one question.

CHAIR BROTMAN: Sure.

CHAIR SEPTIMUS: Just from the call I am assuming that you wrote because the measure is imprecise, it is not valid, do I assume that based on the fact we are trying to be all-inclusive that that took away that comment?

MEMBER SPACH: Yes. Yes, that was the summary from the call.

CHAIR SEPTIMUS: Okay, I just wanted to make sure because that is what it sounded like.

MEMBER SPACH: Yes.

CHAIR SEPTIMUS: Okay.

CHAIR BROTMAN: All right, let's go to a vote on validity.

MS. KAHN: Voting on 2b, validity; high, moderate, low, or insufficient evidence. You can start.

(Pause.)

MS. KAHN: There is zero high; 17 moderate; two low; and one insufficient evidence.

CHAIR BROTMAN: Okay, so that passes. Let's go to usability.

MEMBER SPACH: Usability, this is a measure that has been in place since 2008. It is a -- but there really weren't any major concerns about usability.

CHAIR BROTMAN: Any discussion? Peter.

MEMBER HAVENS: Is there a mechanism through which concerns about usability could reasonably be expected to be collected?

MS. WINKLER: We solicit input and feedback at any time. We specifically solicit issues around implementation and use of the measure at the beginning of each of these projects. So if you have got any other suggestions on how we might get that feedback, we are all ears.

MEMBER HAVENS: Oh no, ma'am, I didn't --

(Laughter.)

MEMBER HAVENS: I didn't have a suggestion. I was just trying to understand if the absence of data suggested the absence of a problem. This is the absence of reported problems, given a reasonable reporting structure.

CHAIR BROTMAN: All right, any other discussion on this point? Let's vote on usability at this point.

MS. KAHN: Voting on usability; high, moderate, low, or insufficient information. So you can start.

(Pause.)

MS. KAHN: So we have three high, 16 moderate, zero low, and one insufficient information.

CHAIR BROTMAN: Okay, let's move on to feasibility.

MEMBER SPACH: The feasibility by our subgroup was viewed to be high. Well actually there was two votes for high and two in the medium, and there was no concerns that we had about the feasibility. This is part of regular medical care and the feasibility of extracting we didn't have any concerns, unless somebody else in the group remembers it any other way.

CHAIR BROTMAN: Aaron, go ahead.

MEMBER MILSTONE: So just to clear it, I think is my question from before and this comes to feasibility.

So you mentioned exceptions to why testing wouldn't be done within 12 weeks. I'm not sure if -- I don't know personally what those are. And if there are exceptions, how are they captured and would it make it harder for other places to capture those that aren't recording them in their EMR or something? So that is in the denominator exclusion. It doesn't give examples. It says exceptions may include medical reasons and patient reasons. Then in your flow diagram it has under -- it has boxes for medical exemption and patient exemption. I'm just trying to figure out what those are and if other groups for feasibility will be able to identify those exceptions to keep them out of the denominator.

MEMBER SPACH: I know one of the concerns was that a client who may not show up in that time period or have a scheduled test

to come and not turn up for it.

CHAIR BROTMAN: Peter?

MEMBER HAVENS: This is a question about the data to try to understand if I am looking at the data that are appropriate for answering this question. In 2b5.1 where this says that there were 83 professionals who were asked to report in this CMS quality reporting initiative, I think, 48 percent of professionals satisfactorily reported. Does that mean -- is that a measure that is hard for 52 percent to do? I just don't understand but I can see my friend back there laughing at me, which I don't take personally.

But I don't -- is that the right data point I am supposed to be looking at to understand if this is hard to do or feasible in a practice setting? I just -- it is my first time here.

MS. BOSSLEY: You know, I think they are having a hard time hearing you. The air conditioner, there is a big blower -- the blower

is going in the back.

MEMBER HAVENS: Oh, I'm sorry. If you look at 2b5.1 where it says that, is that a data point that I can use to try to understand the feasibility or lack of feasibility inherent in this reporting structure?

Part of the issue here is that these things make intrinsic sense to report but the data quality is very low in many of these circumstances and so if we are going to satisfy the requirements of NQF certification, we have to make sure that we understand what data are being brought to bear to answer these questions. So just trying to get to that.

MS. CHRISTENSEN: Yes. So this information that is presented there is from the PQRS program. As was mentioned before, the PQRS program is not administered by us. There are evident challenges with reporting to PQRS that may or may not have to do with internal properties of the measure itself. Not a good answer but, unfortunately, we are not able to

access the raw data to be able to provide any information about where the challenges specifically are.

I'm just trying to see where the decimal point is there. The problems can have to do with a lot of different things but typically it is in how the codes are submitted and whether or not they are submitted with the right QCD combinations.

MS. BOSSLEY: So just to clarify, what you are seeing here really talks about how well they did at comparing the codes that are needed for PQRS for this measure on a claim. This has nothing -- so often what will happen is they submit a code with something that actually doesn't match the denominator. And so it shows it is not satisfactorily reported. It doesn't have anything to do with the performance of the measure.

MEMBER HAVENS: No, it has to do with the feasibility of using this in an electronic way.

MS. BOSSLEY: But that is the data, yes. This is claims. This is using claims data.

MEMBER HAVENS: Oh, okay. So this we are not using --

MS. BOSSLEY: This is not, this isn't EHR-based. This is PQRS claims.

MEMBER HAVENS: Thank you.

CHAIR BROTMAN: Okay, if there is no more discussion, let's vote on feasibility.

MS. KAHN: We are voting on feasibility; high, moderate, low, or insufficient information. You can start.

(Pause.)

MS. KAHN: Can we have everyone press it one more time?

(Pause.)

MS. KAHN: We have one high, 18 moderate, zero low, and one insufficient.

CHAIR BROTMAN: Okay, let's go and vote for suitability for endorsement.

MS. KAHN: Looking at overall

suitability for endorsement. Does the measure meet NQF criteria for endorsement; yes or no?

You may begin.

(Pause.)

MS. KAHN: So we have 19 yes and one no.

CHAIR SEPTIMUS: Okay, well this measure will pass. Now we would like to try to get past these last two measures, which are both counseling-like measures before we break for the evening.

So I think Steve, you have you and this is also a PCPI thing. This is hepatitis C counseling regarding the use of contraception prior to antiviral treatment.

CHAIR BROTMAN: Right, and this is another process maintenance measure from 2008. And it is the description is the percentage of female patients aged 18 to 44 years and all men aged 18 years and older with a diagnosis of hepatitis C who are receiving antiviral treatment, who were counseled regarding

contraception prior to the initiation of antiviral treatment. And this is based on administrative claims and also electronic health records.

Concerning impact, I will just give you a heads up, this looked like it was a pure, to our group discussion, a check the box measure type of measure that was submitted. Did the patient receive counseling or not? And it wasn't clear to us, we had quite a discussion, whether the impact would be reasonable. It was not clear how contraceptive counseling actually reduces pregnancy while on ribavirin and especially it is not clear why men with hep C needed to be counseled. I believe PCPI submitted additional evidence leading to a discussion on that but the impact is obviously affecting this population affected but it wasn't sure exactly. The impact discussion presented just defaults to the impact of hepatitis C disease and didn't specifically address the impact of the measure itself.

CHAIR SEPTIMUS: Yes, go ahead.

CHAIR BROTMAN: Yes, Mohamad?

MEMBER FAKIH: I am thinking it is related to ribavirin we are talking here with hepatitis C. And looking at all the drugs that we use in the country, you know, whether they are warfarin or TNF inhibitors, you know, any drug we use have side effects.

And does this have to be a measure? I mean the first question would be is how would we have -- do we have to go that deep into measures? Because there thousands of drugs that have pretty much maybe even use heterogenicity that we don't have measures for.

CHAIR BROTMAN: That was part of our discussion. I mean, if you have a measure just for checking "the box" on this, you could have a measure for every check the box type of issue.

CHAIR SEPTIMUS: Let me ask John in terms of impact, do we have data that show teratogenicity in patients treated with hep C with infant morbidity?

DR. WONG: So as you know, the label for the FDA approval for the drug makes it, in essence, the equivalent to thalidomide. It is the class X. So there are very few individuals, women who have become pregnant while taking ribavirin. And secondly, there are very few women who have had male partners who were taking ribavirin at the time they became pregnant.

So I think it is kind of a catch 22 in that we don't really know the teratogenicity in patients or we have a very small sample of that. It is based, as is most of these studies, on animal evidence and there is, again, a dose and duration effect observed in hamsters and rats. Yes, rats and hamsters.

You know, rabbits, there is a mortality but there is no teratogenicity observed.

CHAIR BROTMAN: And no human evidence?

DR. WONG: I provide in the supplemental a bit of human evidence. It is

fairly limited. Let's see. When I put together two studies out of 25 in the literature involving men whose partners became pregnant, 12 normal babies out of the 25, five miscarriages, two elective abortions, seven patients were lost to follow-up.

There is an ongoing ribavirin pregnancy registry which reported their data from 2003 to 2009. They enrolled 49 live births with direct exposure to the mother and 69 live births with indirect exposure based on the male partner. They found six birth defects in those pregnancies.

CHAIR SEPTIMUS: So how would that compare to the general population?

DR. WONG: Yes, I don't know the numbers off the top of my head for the general population.

CHAIR BROTMAN: I hate to put my epidemiology hat on but --

CHAIR SEPTIMUS: Tom, go ahead and comment.

MEMBER FILE: Well my point was I mean all these people had hepatitis C so you would want to compare it with those people which probably you don't have a lot of data either but not just the normal population.

CHAIR SEPTIMUS: So we may want to have -- oh, Tom. Let me get Tom first.

MEMBER GIORDANO: So I think it is reasonable to expect that ribavirin is teratogenic in this population. There is no reason to expect that it wouldn't be. But the question is how many women get pregnant while they are on hep C therapy and how many men impregnated women while they were getting hep C data.

Are there any data on that, which is the impact?

DR. WONG: So I don't know except for those reported in the literature. I do know that the initial report by Willis Maddrey was based on some of the randomized controlled trial data. So, even despite all of the attention

that gets placed on them during the RCT, where they actually were supposed to get monthly pregnancy tests, some of those women did go ahead and get pregnant.

Now how often this happens and does it not happen because of counseling or black box warning, I don't know.

CHAIR SEPTIMUS: And the other question I think for us is do we now that counseling has a high impact on this adverse event. So as I read this, you know, does this meet the NQF description of high impact? And I think that is what we need to decide on right out of the box.

So Aaron.

MEMBER MILSTONE: I was just going to add I think it is also even one step farther back. The outcome isn't pregnancy it is birth defect. So does counseling prevent pregnancy, which then has an impact on birth defect.

CHAIR SEPTIMUS: That's a very important point.

Any other question before we vote on this? Then let's vote on high impact.

MS. KAHN: Voting on 1a, high impact; high, moderate, low, or insufficient evidence. You can go ahead and start.

(Pause.)

MS. KAHN: So we have zero high; two moderate; four low; and 13 insufficient evidence.

CHAIR SEPTIMUS: Okay, well this is a stop, as you know, for this measure. So this measure fails.

So let's go on to the last measure of the day, which again has to do with counseling regarding alcohol consumption. I think that is your, Mary.

MEMBER BLANK: Okay. Again, this is a process maintenance measure approved, I believe, back in 2008, also assessed through administrative claims, EHRs, electronic clinical data, and registries. In regard to the impact when the working group -- well first

of all we heard Dr. Wong provide an epidemiological overview of the extent of hepatitis C and the extent of the illness and the severity of it, the statement that hepatitis C virus infected individuals with high alcohol intake have more severe fibrosis, more rapid progression and a higher rate of cirrhosis and hepatocellular cancer.

There were eight citations listed, two of which discuss alcohol impact on hepatitis C infected individuals and the working group felt that the hepatitis C, the information that was conveyed that the measure addresses counseling for alcohol consumption but it does not equate to cessation and it is going back to the document recommendation about avoiding recommendation that can be a sort of check the box type of documentation.

Did I miss anything from those of you in the working group that you wanted to add in regard to impact?

CHAIR SEPTIMUS: So like the other

measure, I will ask John, do we know that the impact of counseling on alcohol consumption?

DR. WONG: So there are smaller studies within the hepatitis C infected patients of brief interventions. The larger body of data was obtained in two systematic reviews, one demonstrating modest effect and the other one was more specific in terms quantifying the reduction. This was in a primary care setting with a brief alcohol intervention and they demonstrated a reduction of somewhere between two to five drinks per week, based on 19 randomized controlled trials with 5600 patients.

CHAIR SEPTIMUS: Tiffany?

MEMBER OSBORN: Do we have information about whether or not that was sustained?

DR. WONG: No. Typically those trials are for a relatively delimited time, as is most studies for cost reasons.

MEMBER OSBORN: Sure. So what was

the general time frame that we were talking about that this reduction was effective?

DR. WONG: Usually they go out to six months to one year.

CHAIR SEPTIMUS: Kathleen?

MEMBER BRADY: What were the specifics of that intervention? What did that involve?

DR. WONG: They are described as something that you would do in the course of normal counseling with a patient. So it is a brief interaction with the patient about the relative harms of alcohol.

I will mention that there is a paucity of data among patients who are heavy drinkers or dependent drinkers. So there were 16 RCTs in all. Out of those 16, 14 excluded heavy drinkers or dependent drinkers, again, because they anticipated that a brief intervention would have very little impact on those patients.

CHAIR SEPTIMUS: Doug?

MEMBER CAMPOS-OUTCALT: I just have a question. So if undergo treatment and you get sustained viral release, can you go back to drinking?

CHAIR SEPTIMUS: That's not a personal question is it?

(Laughter.)

DR. WONG: Am I your doctor?

MEMBER CAMPOS-OUTCALT: I don't have chronic hepatitis C.

DR. WONG: So our data suggests that once you get rid of your hepatitis C, fibrosis that you have in your liver tends to resolve.

So the only question in my mind is we tend to see a lot of patients now who have both hepatitis C and fatty liver disease. And in the presence of fatty liver disease, we would discourage it.

CHAIR SEPTIMUS: Adam?

MEMBER THOMPSON: I have a question for you. As far as the study you are studying about, the behavior intervention that was done,

was that conducted by the physician or was that conducted by a non-medical professional like a case manager?

MEMBER SPACH: I don't recall the specifics in all of the RCTs. I suspect, I could be wrong, that there is a mixture of both.

MEMBER THOMPSON: Were you all, when you define it in your numerator as far as who received that counseling, were you meaning to specify that as any type of person or did it specifically mean the person's physician?

DR. WONG: It does not specify the physician. It simply has to be documented in the chart so that if it is a physician extender, a PA, a nurse practitioner, somebody who counseled the patient, that would be adequate.

CHAIR SEPTIMUS: Tom? Okay. Mohamad.

MEMBER FAKIH: You know, this is the issue with documentation does not mean that we really counseled. It may have been okay, don't drink or sitting down for 20 minutes with a

patient and this is not clear also.

DR. WONG: It is one of those issues with counseling that you all have sort of mentioned. You know, short of documenting the quality, extent, the coverage, having a patient sign something that you did this.

I will say that roughly about half, in terms of performance gap that we will get to eventually, roughly about half of patients have this documented. And again, there are the question that you all raised.

CHAIR SEPTIMUS: Mike, go ahead.

MEMBER FARBER: Well my comment would be is I don't at all think that there is any reason not to counsel many people about drinking. We have people that are on all sorts of diseases and drugs and treatment of which alcohol would interfere with it. My question here is is this a necessary measurement for this particular issue? Is it, in a sense, if you were on Antabuse, you would definitely absolutely want to counsel someone. And the

results of drinking would absolutely be definite.

So I guess that is what I wonder. Is this too much of a burden for providers as another measurement?

DR. WONG: So I think that is a great question or comment. Two things. One is the CDC guideline that just came out. Again, if you are not thinking about antiviral treatment, one of the key prognostic elements for progression is alcohol intake. And that was demonstrated by Terry Poiniard in a Lancet study suggesting that individuals with hepatitis C who drank five or more glasses of alcohol in whichever form you prefer, progressed much more rapidly than those who did not.

CHAIR SEPTIMUS: Peter?

MEMBER HAVENS: So the question about impact versus data demonstrating impact becomes crucial here as it might have been previously. In my mind and in the mind of the CDC, this represents the standard of care in

anyone with hepatitis C you tell them if you drink, your disease will progress more rapidly. So you should not drink.

Now we are arguing about the potential benefit of that. At the beginning the staff led with this don't pass a check the box or somebody is going to get mad at us. And this is a kind of a check the box thing on the one hand. On the other hand, it is the standard for most organization. So I am sort of at sea about how to vote.

MS. WINKLER: I think just the guidance, it actually talked about measurement. If you remember, the second bullet was teaching and counseling should be assessed from the patient's perspective how well were you counseled or were you counseled or something along that line to determine the effectiveness of the counseling, as opposed to a more check the box somebody said something.

CHAIR SEPTIMUS: And the question -- there seems to be clearly an opportunity,

which we haven't gotten to yet in the gap, but the question is is this particular measurement that could be a check the box. It could be like smoking cessation, which most of us have been involved in in a while but the question is, is it a high impact in terms of changing behavior. And that is something that we might not be able to answer.

Tom?

MEMBER GIORDANO: So I actually am very accepting of the fact that brief counseling from a physician on the topic of alcohol intake is impactful. It may not be a 50 percent reduction or impact 50 percent of the population, but there is pretty convincing data that it will work, it will reduce alcohol intake in a reasonable percentage of the patients. That may only be five, ten, 15 percent, but for a two minute intervention, that is a pretty good bang for your buck.

So I am willing to accept that this is a population that is a big population, alcohol

is a problem in this population and there is an intervention that does work, according to many, many randomized controlled trials.

My question is just whether we want these check the box measures but that is a separate issue.

CHAIR SEPTIMUS: Kathleen. Mohamad then Kathleen. Excuse me.

MEMBER FAKIH: You know, I think it is how we are framing this question. Do we have a high impact having this measurement of just documentation versus was actual counseling done and how we assess that. And this is the trouble I am having.

CHAIR SEPTIMUS: Kathleen?

MEMBER BRADY: I was going to make the same comment that there is no clear definition of what counseling means. And that is a big problem for me. And it is going to vary based on one person to the next.

CHAIR SEPTIMUS: Tiffany?

MEMBER OSBORN: So my question is

you quoted data in patients that did not have hepatitis C, right? And we were hoping to extrapolate. But then you also provided the qualifier that they excluded heavy drinkers.

So I do wonder whether or not the population that we are extrapolating this to is actually excluded from the studies. So do we really have any data?

DR. WONG: So in the hepatitis C patients in the CDC guidance that just came out about ten days ago, about 58 percent of patients with hepatitis C drank two or more alcoholic drinks per day. So again, that doesn't give me the tail which are the heavy dependent ones, but again two or more than that would fit within my range of patients who ought to be counseled.

CHAIR SEPTIMUS: Any other comments about high impact? I'm sorry, Adam.

MEMBER THOMPSON: I just wanted to back up what Tom was saying about it being delivered by a physician. But to draw everyone's attention to the fact that this is

not written as a physician-delivered intervention specifically.

And if it was, I would be more likely to agree with it. But because this could be delivered by anyone in that clinic, I am less likely to see it as a meaningful outcome because I have seen how it can be delivered inappropriately.

CHAIR SEPTIMUS: Michael?

Well, let me ask you, Adam, would you accept a physician extender? Because sometimes they actually spend more time with patients than the physician.

MEMBER THOMPSON: I think there is something to do with the fact of your doctor specifically taking the time to speak with you about something that is behavioral, that has more of an impact than a nurse, a case manager, or even a peer.

I think it can be supported by an extender but I think as far as the initial conversation there is major impact of a doctor

taking that time.

CHAIR SEPTIMUS: Yes, Doug.

MEMBER CAMPOS-OUTCALT: I don't want to be contrary but the evidence doesn't support that. The evidence is pretty good that physicians -- an initial statement might be beneficial but physician counseling is not as effective as other people who have been trained to counsel. And those people who are actually trained to do that do a better job than physicians.

MEMBER GIORDANO: Yes, I would like to second that as well.

CHAIR SEPTIMUS: Tom has been arisen here.

MEMBER GIORDANO: Sorry. I think both sides of this are right. Yes, a trained counselor is very effective at decreasing alcohol use. A brief physician message is effective as well but I think what Adam was talking about was an RN who is taking your vitals delivering the message or a med tech who is

checking you in delivering. There, I agree that there is no data to suggest that that is effective. So I think both comments are accurate.

CHAIR SEPTIMUS: Okay, if there is no other comments, I think it is time to vote on whether or not -- oh, I'm sorry. I didn't see you. I apologize.

MEMBER RAMIAH: That's okay. So I am a trained smoking cessation counselor. And comparing smoking cessation counseling to alcohol counseling, if there are steps laid out, it is feasible. Since smoking cessation counseling you have the five As that is laid out and that is what the physician or physician extenders or anybody in the practice could do. So that is my missing piece in this impact.

DR. WONG: So there are a variety of interventions that are available within the literature and there has not been the unification that has occurred within smoking cessation.

I think, Katherine, behind you wanted to make a comment.

MS. AST: Really quickly I just wanted to also submit that we have another measure in our suite of measures that is an alcohol screening and brief intervention measure that has a definition of brief counseling. So it is possible for us to take that definition back to the workgroup and see if we can incorporate it into this measure.

So I don't have it in front of me. I'm sorry about that but it specifies about five to 15 minutes and it talks about what are some possible things that could be discussed.

CHAIR SEPTIMUS: Thank you. Did I miss anybody again? I hope not. I apologize if I looked the wrong way.

So why don't we go ahead and vote then on high impact?

MS. KAHN: Voting on 1a, high impact; high, moderate, low, or insufficient evidence. You can go ahead and start.

(Pause.)

MEMBER COLLINS: There is two that are missing.

MS. KAHN: We have one high, five moderate, six low, and six insufficient evidence.

CHAIR SEPTIMUS: The measure fails and so I guess we shortened our day by about -- well I know from this group you really stayed engaged and really in it the whole time. It is incredible to go from 8:30 until 6:00 and have the kind of razor sharp comments. So I know that Steve, I'm sure, feels the same but he can speak for himself. A great discussion today.

CHAIR BROTMAN: I think if we went any further the cookie man would revisit us.

MS. WINKLER: Okay just a couple follow-up things. I mean, earlier when we were talking about the sepsis measures one of the votes was on insufficient information and I think that NQF we feel like we kind of let you

down, particularly since that data collection tool wasn't in your materials although it should have been. So we actually want to give you a copy of that to talk a look at that.

I think we have got four measures that we had hoped to do today that we are going to have to add on tomorrow's schedule. And as you can see, this takes a while to go through this iterative process. Hopefully, you know, everybody is a little bit more tuned in and we can be focused on getting through them because I know you are going to start having planes to catch, you know, come around 2:00, 2:30 in the afternoon.

So does anybody have an objection to starting at 8:00 in the morning? We are scheduled for 8:30. That would move it back 30 minutes. That would be 7:30 for breakfast.

CHAIR SEPTIMUS: So I know we can do it. And how many drinks can we have tonight, John?

DR. WONG: I would have to know some

protected medical information first.

CHAIR SEPTIMUS: Okay, before we go, operator, if you would open up the lines and see if there is any public comment before we officially adjourn for the day. Or any in the room also. Excuse me.

OPERATOR: If you would like to ask a question or have a comment, please press *1 on your telephone keypad.

There are no questions or comments from the phone line.

CHAIR SEPTIMUS: Well thank you very much and we are officially adjourned for the day.

(Whereupon, the above-entitled matter went off the record at 5:55 p.m.)

1