NATIONAL QUALITY FORUM

+ + + + +

GASTROINTESTINAL/GENITOURINARY ENDORSEMENT MAINTENANCE STEERING COMMITTEE MEETING

+ + + + +

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 8:30 a.m., Andrew Baskin, MD and Christopher Saigal, MD, Co-Chairs, presiding.

PRESENT:

ANDREW BASKIN, MD, Aetna, Co-Chair
CHRISTOPHER SAIGAL, MD, UCLA Medical Center, Co-Chair
LILIANA BORDEIANOU, MD, Massachusetts General Hospital

ZAHID BUTT, MD, Medisolv, Inc.
ROBERT ELLIS, Consumers' Checkbook
NANCY FALLER, RN, MSN, PhD, CWOCN, Nursing for Wellness
ED GILL, MD, Virginia Commonwealth University Medical Center
JOHANNES KOCH, MD, Virginia Mason Medical Center
JENIFER LIGHTDALE, MD, MPH, Children's Hospital Boston
ALAYNE MARKLAND, DO, MSc, University of Alabama at Birmingham
PAUL MERGUERIAN, MD, MS, Seattle Children's Hospital

JOHN MORTON, MD, MPH, Stanford University

ANNE PELLETIER-CAMERON, MD, University of Michigan Hospitals & Health Centers

STUART REYNOLDS, MD, MPH, Vanderbilt University Medical Center

PHILIP SCHOENFELD, MD, VA Ann Arbor Medical Center

JUDITH TOBIN, PT, MBA, Centers for Medicare & Medicaid Services

NQF STAFF:

TAROON AMIN, MA, MPH

HEIDI BOSSLEY, MSN, MBA

ANN HAMMERSMITH, JD

KAREN PACE, PhD, MSN

ASHLIE WILBON, RN, MPH

EVAN WILLIAMSON, MS, MPH

T-A-B-L-E O-F C-O-N-T-E-N-T-S

T-A-B-L-E O-F C-O-N-T-E-N-T-S
(Continued)

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

(8:30 a.m.)

CO-CHAIR BASKIN: Good morning. We're going to start on time. I think within one minute is on time. I hope everybody had a restful night. We should have an exciting morning. I guess we just got a hint of the GI measures at the end of the day yesterday, but today will be GI day.

I don't have too much to say about a recap for yesterday, other than it was a learning experience. I think we started to find our place in terms of what we consider evidence submitted, evidence implied, evidence probably-out-there, and I'm proud to say we didn't use the exception rule very often, right? I think we stayed within --

(Laughter.)

CO-CHAIR BASKIN: There's a fear that we jump into that too often by everybody, and I'm glad we didn't.

There's been a request that we

change the order of reviewing the measures, so when we do the first measures, the C2056 we're going to do at the end of the 8:45 session, so we're going to do 658 and 659. I think they will be easier to start with, and it will help us when we get to C2056 if we've gone through those first.

Do you have anything in particular you want to say, Chris, this morning?

CO-CHAIR SAIGAL: No. Let's just get it started.

CO-CHAIR BASKIN: All right. Then we're going to start with 658. And Johannes, you're up -- whoops. See that? I already made a mistake. Developer, the developer. So is the representative for the AMA-PCPI here? And that's both of the 0658 and the 0659, so if you could just spend three minutes or so talking about both measures, not just one, introducing them? Thank you.

(Interruption from phone.)

DR. PARK: Good morning. I'm

Walter Park, a gastroenterologist representing the American Society for Gastrointestinal Endoscopy. I'm here with Maged Rizk, who is here representing the American College of Gastroenterology. Joel Brill, representing the American Gastroenterological Association, is on the telephone. And together, our three societies represent virtually all practicing gastroenterologists in the United States.

On behalf of the PCPI, the three GI societies just mentioned, and a multi-stakeholder workgroup, I would like to briefly introduce two measures for your consideration: 0658 and 0659. Both address the appropriate timing of follow-up colonoscopy. 0658 addresses the appropriate follow-up intervals for colonoscopy in average-risk patients with a normal examination, and 0659 addresses colonoscopy intervals for patients with a history of adenomatous polyps.

The intent of both measures is to

avoid inappropriate use or overuse of colonoscopy. These measures were developed in 2008. The multi-stakeholder workgroup was chaired by Doctors John Allen and Douglas Faigel. I believe Dr. Faigel is with us via telephone today, as well.

These measures received time-limited NQF endorsement in 2011. Testing data was submitted earlier this year for 0658, and testing data for 0659 will begin later this year.

Regarding the importance of these measures, colorectal cancer is the third most common cancer and the second leading cause of cancer death in the United States. The vast majority of colorectal cancers arise from adenomatous colon polyps. The progression from polyp to cancer occurs over an estimated five to ten years in average-risk populations. Finding and removing polyps during this window interrupts malignant transformation and reduces the incidence of, and mortality from,

colorectal cancer.

There are numerous studies to support these measures, beginning with the landmark National Polyp Study published in 1993, that demonstrated that patients who underwent colonoscopy and had polyps removed developed colorectal cancer up to 90 percent less than untreated historical controls. Colonoscopy is considered to be the most effective screening option for colorectal cancer. This procedure directly visualizes the entire extent of the colon and rectum, and permits immediate polypectomy and removal of macroscopically abnormal tissue.

The timing of follow-up colonoscopies should be tailored to the number, size, and pathologic findings of adenomatous polyps removed. For average-risk patients with a normal exam, colonoscopy is recommended approximately every 10 years in all current guidelines. Current guidelines recommend that patients with one to two small

tubular adenomas, defined as less than one centimeter, with only low-grade dysplasia, should undergo follow-up colonoscopy no earlier than five years. Patients with advanced adenomatous lesions, or greater than three adenomas, should have a repeat colonoscopy in three years. A shorter interval of follow-up is recommended in those patients with numerous adenomatous polyps, and in patients with large cecal adenomatous lesions where complete removal is uncertain.

These guidelines assume a complete examination, a high-quality bowel preparation, and complete removal of all visualized polyps. When these assumptions are not met, it is appropriate to reschedule colonoscopy within one to two months to ensure a high-quality examination.

After a normal surveillance colonoscopy, repeat examinations should be done at five year intervals. Performing colonoscopy too often not only increases

patients' exposures to procedural harm, but also drains limited resources that could be more effectively used to adequately screen those in need.

The evidence for these measures has recently been revisited and updated by the U.S. Multi-Society Task Force on Colorectal Cancer. Released electronically this past July, they found that the growing evidence continues to support these measures.

Despite strong evidence for these measures, a performance gap exists, providing an opportunity for improvement. In a 2006 study of over 1,200 colonoscopy reports, recommendations were consistent with the current guidelines in only 37 percent of cases. Further, the adjusted mean number of years in which repeat colonoscopy was recommended was 7.8 years following a normal colonoscopy.

Four different surveys have indicated that post-polypectomy surveillance

colonoscopy in the United States is frequently performed at intervals than those that are recommended in guidelines.

In closing, we want to thank the Steering Committee for considering continued endorsement of these important questions, and are available for questions and clarifications. Thank you.

CO-CHAIR BASKIN: Thank you very much. Does anybody have a particular question to the presenter before we proceed with our review? Zahid, go ahead.

MEMBER BUTT: The submission states that it will be a stage two with electronic specification available. When will that be? It's 0658, there's mention in the submission that there will be a stage two, an electronic specification will be available in stage two.

MS. AST: Just as we stated, it will be ready for the measure submissions when they're due. I believe that might be

December.

MEMBER BUTT: Okay. Thank you.

MEMBER BORDEIANOU: As far as the issue of impact --

CO-CHAIR BASKIN: Well, we haven't presented the case yet. So just if there was a question for the developer, that was that part there.

Johannes, then, let's go ahead. We'll go with impact first.

MEMBER KOCH: All right. So this is a process and overuse measure. I think that it was quite clearly stated that this has potential large impact, given the number of colonoscopies currently being done for screening in the United States, as well as colon cancer being a very prominent and common condition. So I think that the rationale for this being a high impact is pretty well stated, in my opinion.

CO-CHAIR BASKIN: Any comments or questions from anybody regarding that?

MEMBER SCHOENFELD: I would only add that the ABIM and Consumer Reports have come out with their "Choose Wisely" campaign. Many of you may have seen the big displays in the malls about this, about overused medical procedures that patients should question their physicians about when ordered. And screening colonoscopy done sooner than 10 years, and surveillance colonoscopy for polyps done sooner than three to five years are actually the two big GI ones.

So there's a lot of literature about it, but it's also becoming a bigger public health issue in terms of publicity.

CO-CHAIR BASKIN: I think this is probably not terribly controversial, so we could probably vote on the impact quickly here. Oh, we're not using voting buttons today. That's right. Because our ID guys usurped them.

(Laughter.)

CO-CHAIR BASKIN: Okay. So I

guess we're going to vote. You'll be able to raise your hand. I'm going to ask for a 1, a 2, a 3, or a 4, right? Because we have four options here. So think for a second about which one you're going to do, because once I get past one to two, it's too late to go back to one, okay?

(Laughter.)

CO-CHAIR BASKIN: Now, I know this is difficult. All right. So we're going to have a count for each one. One is high, so who's voting for high impact? And who's counting?

(Show of hands.)

MR. AMIN: It's unanimous. 15 high.

CO-CHAIR BASKIN: Well, then, there should be no twos, threes, or fours. Just to make sure, no twos, threes or fours? We only have 15, right?

(No response.)

CO-CHAIR BASKIN: Okay.

MR. AMIN: That's 15 high.

CO-CHAIR BASKIN: Yes, that was 15 high. Yes. All right. Then we'll move on to the evidence base. Johannes, if you want to go ahead again?

MEMBER KOCH: So there's a large body of evidence that's cited. Unfortunately, as in so many other cases, the developers don't actually grade the evidence. I think that there's an overwhelming body of evidence. None of it is high quality, no randomized clinical control trials, but a large number of series, started with the National Polyp Study that was cited.

So there's a large quantity. The quality is moderate, and it's all very consistent, both supporting the use of colonoscopy as well as the fact that there needs to be proper intervals associated with that.

CO-CHAIR SAIGAL: A question, Johannes. Is there evidence that 10 years is

not overuse?

MEMBER KOCH: No. Again, the evidence is that 10 years is -- there's nothing that suggests that 15 or 20 years might not be better or just as good. So 10 years is kind of the lowest bar.

CO-CHAIR SAIGAL: Is there an expert consensus, then, that 10 years is the number?

MEMBER KOCH: Well, I think that there's some evidence from the National Polyp Study that when you look at patients, the interval -- the timing of colonoscopy, the protective effect of it diminishes over time. So I think it's more than just consensus, but I don't think it's locked in time, and I don't think it's been studied beyond 10 years.

MEMBER SCHOENFELD: I would note, there is evidence from prospective cross-sectional studies that five years is too soon. There's not evidence from prospective cross-sectional studies that 10 years is the

right time, say versus 15 years.

Having said that, the available natural history data, which does date back to the 1960s, where you did serial barium enemas to assess the growth of polyps, because it was a choice of either doing a surgical resection or leaving it in place, shows that the average time for an approximately one centimeter polyp to develop into a cancer is between five and 10 years. So that's part of the expert opinion that led to that choice.

CO-CHAIR SAIGAL: So that's in the document?

MEMBER SCHOENFELD: That's certainly in the guidelines that are cited.

CO-CHAIR BASKIN: Any other comments regarding the evidence?

(No response.)

CO-CHAIR BASKIN: Then I think we're ready to go to a vote for this evidence. You're saying that there's a significant body of evidence, some limitations on the upper

limit of study, but there's some reasonable science behind that number of 10 years.

DR. PACE: I just want to point out to everybody that this is a measure, not measuring the interval, but measuring that a recommendation was made. So, just so you know.

CO-CHAIR BASKIN: That's a later comment I have. But as is.

Okay. Well, our choices here are one, two and three. So one is yes, a body -- so again, think ahead. Two: the evidence does not meet it. Three: that it was insufficient in terms of what was presented.

I think you've made it clear, though, there was significant evidence presented, so the question is whether it's good or bad, I guess. But everyone has an option for three. So, let's go.

Option 1. Yes, the body of evidence meets the guidance.

(Show of hands.)

CO-CHAIR BASKIN: Fifteen. Okay, so there are no twos and threes, then. All right. That's straightforward. Let's move on to performance gap.

MEMBER KOCH: So again, I think that it was well-stated, and it's well-documented in the literature as well, that the performance gap here is that there's overuse, that people are doing colonoscopy too frequently, and that the guidelines that are outlined both here and in the literature aren't being adequately followed, so that as a performance measure of overuse, I think this is a very worthwhile first step.

MEMBER BORDEIANOU: I know we're saying that it's a performance measure of overuse, but really the denominator here is this "at least 10 years follow-up." So if you recommended a one month follow up, it would still count as check. So the measure doesn't really measure overuse.

MEMBER KOCH: It's measuring

whether or not you correctly recommended a 10 year interval for somebody who's low-risk. So if you say two years for somebody who is low-risk, you would not meet the quality measure. If you say 10 years, you --

MEMBER BORDEIANOU: That's not what it says here. Recommended follow up of at least 10 years.

CO-CHAIR BASKIN: Of at least 10 years.

MEMBER BORDEIANOU: So it could be a one month follow up.

CO-CHAIR BASKIN: No, no. "At least 10 years" means 10 years or greater.

MEMBER BORDEIANOU: All right. That's why English is my second language. Thank you.

CO-CHAIR BASKIN: So anything less than 10 years would be a non-hit in the numerator.

MEMBER BORDEIANOU: I'll shut up.

CO-CHAIR BASKIN: No, it's good to

ask. Any other comment regarding the performance gap? So you're saying there's reasonable evidence here of a significant performance gap, and I think that's pretty obvious.

Well, then, I think this is pretty straightforward as well for performance gap. So once again we have four choices. Vote one for high. Raise your hands, please.

(Show of hands.)

CO-CHAIR BASKIN: Okay, and that's unanimous, so we'll move on from that. I think then we have one more vote to take, right? Recommending or not. And this is going to go down to the wire, I know it.

(Laughter.)

CO-CHAIR BASKIN: Any discussion prior to this vote? Does anyone want to put forth any particular position?

(No response.)

CO-CHAIR BASKIN: Didn't think so. Okay. So one, raise your hand for yes.

(Show of hands.)

CO-CHAIR BASKIN: Okay. Then the recommendation unanimously was for approval of the concept. At this point, though, there may be some comments for the developers before this moves forward to Stage 2. Any recommendations or thoughts?

I have one particular one that I'd like to bring up, and that is the possibility that the measure is actually backwards, and that we should be looking at colonoscopies that were performed, and look back to see if it was a normal surveillance colonoscopy, was it 10 years or more since the prior colonoscopy?

And the reason I say that is because this is -- just as Karen mentioned, this is a recommendation of an interval of 10 years. And in fact, if you measured it the other way, we'd be measuring the actual outcome. You know, was it performed 10 years or later? Not what was recommended. In 10

years, we don't know when people are going to get their actual colonoscopies. So to me, it's the difference between telling somebody to do something and actually having it done.

Plus, I think it would avoid a lot of the situations where when you do a colonoscopy, you're recommending at the time surveillance in 10 years, should the biopsies be all normal, or whatever. Because there's a lot of biopsies taken that may be a hyperplastic but not an adenomatous polyp, and therefore wouldn't change the screening interval.

And since that information is not known at the time, oftentimes, of dictating the report of the colonoscopy and recommending 10 years -- and in fact you may or may not be making the right recommendation -- if you reverse the measure, you will always know whether the surveillance was truly an average-risk surveillance at the time the next colonoscopy was done.

So that's one thought of a way to consider doing it. Phil?

MEMBER SCHOENFELD: I would agree that actually both need to be performance measures. And although there are logistical issues with determining how to develop a performance measure when polyps are biopsied, what more recent data has demonstrated is that, even when a colonoscopy is normal, that if the quality of the bowel cleansing is not optimal, that up to 40 percent of the time the colonoscopist, at the time of writing the report, even though it's a normal colonoscopy, will say five years or three years to repeat it, because they're worried they might have missed something.

And this is a discussion for another day, but I would just note that both need to be performance measures, because, as a primary care physician in a managed care organization, if the GI doc says "Well, you need to repeat this in five years," and the

patient somehow manages to stay in the same managed care organization for five years, primary care doc's going to send him back at that time.

In addition, if it's been documented as normal colonoscopy, absolutely, you want to be able to document, when you repeat that screening colonoscopy, that you can document that yes, you're doing it appropriately, and in fact now for the next measure, which is colon polyp surveillance, that's exactly what's being recommended.

MEMBER KOCH: So I think that the biggest limitation is that we don't have data from patients previously. I mean, patients change health plans so many times that knowing what their colonoscopy was five or 10 years ago is frequently a very difficult process. Just as a point, though, if you say that the bowel prep was inadequate, you can do it in less than 10, as long as you have a rationale.

Now, I think it should -- my

recommendation would be that the number of times that you are recommending something outside -- you know, if you write "bad bowel prep" in every single patient, you will have met quality guidelines, you'll just have done lots of bad colonoscopies. So I think the number of times that this exception is used actually should be monitored, right? So the rate at which you're going out of guideline is a quality measure as well. Just because you state that you repeated it in five years and have a rationale for it --

MEMBER SCHOENFELD: That's coming up in the third one.

MEMBER MORTON: The only thing I'd say about utility of the bowel prep is, it's not always physician-directed. And if you have different populations, you may have to make different sort of accounting for who's going to be bowel prepped or not.

MEMBER MARKLAND: I'd just like to, as a geriatrician, there's a growing body

of evidence about maybe there's an upper age -- not age alone, but criteria for maybe not doing a colonoscopy in 10 years, and if the developers could somehow consider some of that growing body of evidence as to who may be a good candidate in terms of -- maybe not life expectancy, but multi-morbidity, I think that would be an important piece of a measure like this.

MEMBER LIGHTDALE: The only thing I'd add is, beyond the bowel prep, there are other reasons you might also not be following the guideline -- I like your point, by the way -- including inflammatory bowel disease. Those patients need colonoscopy more often.

CO-CHAIR BASKIN: They do talk about exceptions, exclusions, for above average risk. It may be helpful to define a little more clearly what above average risk may be, other than -- obviously, inadequate prep is not above average risk. But I could see that being an easy way out to do them more

frequently. There's a lot of people who consider above average risk, and what is truly above average risk. That's another story. Okay.

Oh, Zahid?

MEMBER BUTT: I think the exception thing is referenced here in the denominator section. They do say that you should calculate the exceptions as a separate calculation, to track what people are doing.

CO-CHAIR BASKIN: The issue there is whether they're excluded from the denominator or whether they're in the denominator, but you separately calculate the exception in the numerator. I think NQF has been -- at least at the CSAC level -- has been more saying "include them in the denominator, but you can count them as a separate category, exception, in the numerator."

MEMBER BUTT: Yes. There's this whole discussion about the difference between exclusion and exception, exclusion being

excluded from the denominator and exception being that you get credit for it, it stays in the denominator somehow. I think that's kind of how it is coming out to be.

CO-CHAIR BASKIN: Yes. Well, that we can go through, maybe at the next stage. But there's different ways to consider how they count.

Any other comments for the developers before we move on to the next one?

(No response.)

CO-CHAIR BASKIN: Okay. Thank you. The next one, Philip, 0659?

MEMBER SCHOENFELD: So, 0659 refers to the issue that Andy actually mentioned before, which is to say at the time an endoscopist is performing a colonoscopy for colon polyp surveillance -- so, repeating the colonoscopy in someone who's had adenomas identified in the past -- that at that time, they be able to document that they are not overusing colonoscopy by showing that it's

been at least three or more years since their last colonoscopy.

So the specific measure is the percentage of patients over 18 receiving a surveillance colonoscopy with a history of a prior colon polyp who had a follow-up interval of three or more years since their last colonoscopy, with the numerator being the percentage of patients who had that three or more year interval, the denominator being all patients undergoing surveillance colonoscopy, recognizing that there are some exceptions, multiple listed.

For example, if a person all of a sudden had gross hematochezia in that interval, you'd repeat a colonoscopy, although it wouldn't specifically be for surveillance. If a patient had more than 10 adenomas, we normally go back in one year, because of that number.

This is a maintenance indication from the AMA-PCPI. And in terms of impact,

again, similar to what we already discussed for screening colonoscopy. This is considered one of the most overused GI procedures, which is a poor use of health care resources and exposes patients to additional risks. The fact that it's overused is documented in multiple endoscopic database studies, as well as survey studies of physicians, asking them what their actual practice is.

So, I'll stop there for impact.

CO-CHAIR BASKIN: Comments regarding impact?

(No response.)

CO-CHAIR BASKIN: No surprise. Then I think we'll go ahead and vote on impact. So, again: one, two, three or four. So one is high impact. Raise your hands, please?

(Show of hands.)

CO-CHAIR BASKIN: I think we got a unanimous out of that. Okay. Then we'll move on from impact. 15 voted high.

The evidence base?

MEMBER SCHOENFELD: Okay. With respect to evidence, there are multiple randomized controlled trials that have been looked at, both in pooled patient-level analyses, as well as in meta-analyses that demonstrate that performing surveillance colonoscopies sooner than three years in patients with one or more large adenomas, and sooner than five years in patients with more than one or two small adenomas, does not increase your yield for precancerous adenomas, in effect that those are appropriate intervals.

So I would say that the quality of evidence is actually high in terms of coming from randomized controlled trials of high quality, demonstrating that these are appropriate intervals, and they are consistent, and there are more than four. With respect to the fact that there is overuse, that again comes from database

studies, as well as survey studies of physicians.

So they are consistent, and there are more than four. The quality would be moderate, in that they're not randomized, controlled trials.

CO-CHAIR BASKIN: Comments regarding the evidence?

DR. PACE: I just have a question, because that's -- what you're reporting seems to be different than what was in their forms. Are you reporting what you know, or what --

MEMBER SCHOENFELD: With respect to the survey studies, that's actually material that they did not -- well, actually, they did to some extent. If you actually look at the guidelines, those specific studies I commented are subsumed within their guidelines.

DR. PACE: Right. But the meta analysis you talked about, they didn't provide any results. What was the meta analysis on?

MEMBER SCHOENFELD: The meta analysis looked at randomized controlled trials, comparing three year versus five year intervals for patients with one or more adenomas versus more than three adenomas, to determine whether three years versus five years versus shorter intervals was the appropriate interval to repeat the colonoscopy. They actually cited the pool of patient level analysis by Martinez et al. specifically.

But again, the meta analyses are subsumed within the actual guidelines.

CO-CHAIR BASKIN: Other --

DR. PACE: Can I just make one clarification? So, just as we had all these discussions yesterday, what we're asking the developers to do is to summarize the quantity, quality and consistency. And as we talked about, that's definitely an issue with a lot of the guidelines, of actually being able to access that information.

For example, on the meta analysis, having some actual information about what the outcome was that was being studied, and what the effect size was, is kind of what we're looking for, but understand the difficulties with the current guidelines.

CO-CHAIR BASKIN: Other comments regarding the evidence?

(No response.)

CO-CHAIR BASKIN: Okay. It sounds like a little bit of a mixed bag. Some of the evidence was directly submitted, and some of it is behind the guidelines. And while we would certainly prefer that be submitted, at least in this case we have some inkling of what's behind those guidelines.

MEMBER SCHOENFELD: I think I would just say, if -- and maybe this is part of an issue of guidance from NQF to the developer about just how long you want the packet to be. With respect to this topic, probably they could have gone into more

detail, it just would have -- at least within this specific field, considering the volume of evidence, it would have ended up, as they gave you a couple of paragraphs on multiple studies -- would have been longer. But that may be appropriate feedback for the next stage, to include the meta analyses.

DR. PACE: Right. And just to be clear, we're not asking them to summarize every study. It really is in terms of the body of evidence, which can include multiple studies and multiple meta analyses, but not to -- we don't really want a summary of each individual study, because that's not a summary of the body of evidence.

So I realize that it's difficult, yes.

CO-CHAIR BASKIN: Zahid?

MEMBER BUTT: Yes. I think the ideal thing would be to have the guideline, and then whatever studies were used to support the guideline, just listed with whatever links

-- like, some of them have provided links to the studies, if they are available, or at least an abstract that you can link back to. So just a listing of those guidelines behind it, because the guideline itself does not often include the listing of the studies that were used to support the guideline.

CO-CHAIR BASKIN: Other comments?

(No response.)

CO-CHAIR BASKIN: So your vote here is a one, two, or three. One is you're basically saying that enough was submitted and you're comfortable enough that it is there, that's a yes. Obviously two means no, the evidence doesn't meet the guidance. Three is the possibility that it's insufficient, but that it does exist somewhere. That's why we're voting.

So, all those voting number one?

(Show of hands.)

CO-CHAIR BASKIN: And those voting number two?

(Show of hands.)

CO-CHAIR BASKIN: That's zero. And those voting number three?

(Show of hands.)

CO-CHAIR BASKIN: And we have one vote for number three. If I may ask, though, insufficiently submitted, I certainly understand that vote. Do you feel that body of evidence does exist, though?

MEMBER SCHOENFELD: Yes.

CO-CHAIR BASKIN: Okay. A little more comfort knowing that. Then we'll move on to the next part, which I guess is performance gap, if my memory serves me well. Philip?

MEMBER SCHOENFELD: So performance gap for this measure has been difficult to quantify as outlined here, as this goes, I think, more to what might be done in Stage 2, meaning development of better electronic health records, development of better reporting in databases, and better use of ICD-9 codes will allow, for this specific

measure, a better assessment of the performance gap. Meaning at the time you perform your colonoscopy, how well are you documenting that and reporting that yes, it's been three or more years since the last colonoscopy?

So this may be an issue that Karen can provide some further feedback about, but again, in terms of database studies demonstrating that patients who are undergoing colonoscopies for colon polyp surveillance are getting colonoscopies sooner than three years, there's certainly been multiple studies that have demonstrated that. So I guess I'll pause there and ask Karen if she has any other comments specifically.

DR. PACE: Well, one of the things, because this is a measure undergoing maintenance endorsement review, we ask for performance on the measure as specified. And they've reported that the performance rates on this measure, though it's not given by

physician group, I don't believe -- well, maybe it is.

But they say that at the 10th percentile, it's 93.4 percent, and at the 25th percentile it's 100 percent. So this is a measure that really has very high performance rates as it's specified. So the question is, if studies show that there's a performance gap, then this measure is probably not measuring that. So that's part of the issue with an endorsement maintenance, whether the measure as specified is actually indicating that there's a performance gap, or room for improvement.

MEMBER SCHOENFELD: And that's why I wanted to turn it over to you, because I'm not really sure how the measurement is being done, because other published data indicates that there's a much bigger performance gap. And this goes back to at least the way I interpreted their application, talking about how better utilization of electronic health

records would actually facilitate better quantification of what the performance gap is. That's what it seemed to me, but I'm still a novice at looking at these applications, so you might be able to better discuss that issue with relation to performance gaps. Because I did note that it didn't look like the performance gap was that great, based on what they put in here.

MEMBER KOCH: So I think that one of the limitations is that currently much of the data is being gathered through self-selected registries. So the AGA and the ACG have registries to which people voluntarily upload data, to try to become quality leaders in their field.

So obviously, only the people that are going to meet that bar -- you know, you're not going to voluntarily allow your records to be reviewed if you know that you're not meeting the quality metrics. So I think that the difficulty here is that there's a huge

selection bias for what's currently available. Now, applying this as a quality standard outside of self-reporting, obviously, would show much different data, I think.

DR. PACE: Right, but that's not the measure that this is. This is a measure that's self-reported, and these data are from the PQRS program, which is how the measure's been implemented. So that's the issue, is that if you all think that there are substantial performance gaps, then this measure is not doing that. And what is the reason for that?

CO-CHAIR BASKIN: John? I'm sorry.

MEMBER SCHOENFELD: I think the issue is, having been involved in this, the vast majority of GI docs do not report this to PQRS at this time, and that's why initially, in the next two years, PQRS is trying the carrot first. You're going to get a slight increase in your Medicare reimbursement if you

begin to report this. And then in 2015, then the stick comes in, and you actually start getting decreases in your Medicare payment if you don't report this.

So again, going back to what Johannes said, we still have quite a ways to go to improve the number of endoscopists who are actually reporting this. So I'll go ahead and stop there.

MEMBER MORTON: And just for Phil, do you think there's more data out there than what's presented here by the developer in regards to the performance gap?

MEMBER SCHOENFELD: There are certainly multiple database studies that -- these are Medicare database studies that demonstrate that patients who have a colon polyp found are getting their repeat colonoscopy at one or two years. That appears to be overuse based on the review of the Medicare database.

MEMBER BUTT: And I think that the

PQRS program is being aligned with the EHR incentive program, and there is currently a pilot underway. And the alignment between the EHR incentive program and the PQRS, actually, on the ambulatory side, is further along than on the inpatient side. So this probably is going to happen relatively quickly, in terms of being able to generate this measure from EHRs.

MEMBER MERGUERIAN: There may be also a selection bias, because if PQRS takes patients 65 and older, and does not include younger patients, they cite three articles above that that basically say that there probably is a performance gap, up to a 36 percent compliance rate.

CO-CHAIR BASKIN: Any other comments regarding the performance gap?

(No response.)

CO-CHAIR BASKIN: So it seems as if the performance gap is low, at least from the PQRS data, but we've heard comments that

that's not necessarily representative of the gastroenterology population out there. And that makes some sense, and there seems to be some data that's less crystal clear, less superb on the level of quality of data, that there is a gap, but it's unclear exactly how much it is.

So I guess we have to determine -- Oh, I'm sorry. Phil, did you want to make one more comment before we vote?

MEMBER SCHOENFELD: I think part of my question to Karen was this, though. When I read it, it seemed like what was going to happen in Stage 2 of the process is exactly what Zahid was referring to, that in Stage 2 of the process the developers wanted to work out better how to make this measure be correlated better with EHR reporting, so that that way you would get a more precise measurement of this issue.

DR. PACE: I guess we can ask the developer. I mean, basically our criteria are

whether there's a performance gap. And when you have this situation of actual performance on the measure being quite different than you as experts saying is the reality, then the question is, is this actually the right measure to be put forward as a quality performance measure.

But could we ask the developer to indicate whether they're submitting eSpecifications, or if they're going to be submitting the CPT II specifications?

MS. AST: As indicated earlier, we will be submitting eSpecifications.

CO-CHAIR BASKIN: Any other comments before we take this to a vote?

(No response.)

CO-CHAIR BASKIN: Okay. So, once again, we have four choices. How many are voting one, that there's a high performance gap?

(Show of hands.)

CO-CHAIR BASKIN: How many are

voting two, a moderate performance gap?

(Show of hands.)

CO-CHAIR BASKIN: And then how many are voting three, low?

(Show of hands.)

CO-CHAIR BASKIN: None. And how many votes for insufficient evidence, four?

(Show of hands.)

CO-CHAIR BASKIN: One. Did we add up to 15 votes? Yes. So could you?

MR. WILLIAMSON: We have four high, 10 moderate, zero low, and one insufficient.

CO-CHAIR BASKIN: So we've gone through that threshold, so let's go to the next part, recommending the approval of the concept. Is there any additional discussion before we have this vote?

(No response.)

CO-CHAIR BASKIN: Okay. Those voting yes, raise their hands?

(Show of hands.)

MR. WILLIAMSON: We have 15 yes and zero no.

CO-CHAIR BASKIN: And the measure moves on. Any comments for the developers, to help them before this goes to the next stage?

(No response.)

CO-CHAIR BASKIN: No comments? Okay, thank you.

MEMBER BUTT: I have a --

CO-CHAIR BASKIN: Oh, Zahid, please.

MEMBER BUTT: I assume that the measure will be able to make a distinction between whether the recommendation should be three years or five years? Because right now, it says "at least three years," but one of the studies they cite was looking more at the five year number, because of the definition of what it should have been.

In other words, will it be aligned with what the recommendation should be in terms of the interval?

CO-CHAIR BASKIN: The measure is as it is. It's just measuring three or more years. So both those cases are being lumped into "greater than three years."

MEMBER BUTT: Right.

CO-CHAIR BASKIN: It would require a change in the measure to distinguish between three and five.

MEMBER BUTT: They should look into whether another subset of this should be the five years as well.

CO-CHAIR BASKIN: Thank you for those comments. Additional comments for the developers?

(No response.)

CO-CHAIR BASKIN: Okay. Then thank you all. We'll move on to the next measure. So we're going to go back to C2056. Gail, are you still on the phone?

DR. AMUNDSON: I am here, Andy.

CO-CHAIR BASKIN: Oh, great. Gail, you have approximately three minutes or

less. Not three to five, I think we've decided this one measure is three, right?

(Laughter.)

CO-CHAIR BASKIN: If in three minutes, you could just give us a few words about this measure before we do the review? This is the Colonoscopy Quality Index.

DR. AMUNDSON: Yes. This is the Colonoscopy Quality Index --

CO-CHAIR BASKIN: Gail, a little closer to the phone if you can? We're having a little difficulty hearing you.

DR. AMUNDSON: Okay. Is that better?

CO-CHAIR BASKIN: Be quiet and listen closely.

DR. AMUNDSON: Okay. Can you hear me well now, Andy?

CO-CHAIR BASKIN: That's good enough. Thank you.

DR. AMUNDSON: Okay. Colonoscopy Quality Index is a composite all-or-none

measure, and I have been listening to the discussion on these previous measures, and so I think I will be able to short-circuit my comments just a little bit.

But the premise of the measure is that a high-quality screening or surveillance colonoscopy is one that is performed on a patient that needs the test. It is a procedure that's performed in a thorough manner. It is one that is performed without harming the patient.

And so the elements that are included in this measure are whether or not the patient actually needed the procedure they are having today. I would contrast that with existing measures that look at follow-up recommendations exclusively. So in this measure, if a patient is having a procedure at too short an interval, there will be a fail on the composite measure, with the detail providing that the failure was related to the fact that the patient was being screened or

surveilled too soon, or too early.

The other items that are -- the other distinction I would make is that this denominator combines patients who are undergoing both screening and surveillance, so it is a large denominator. It is intended to minimize small number sizes. We have data on individual endoscopists. We've been reporting this measure publicly since 2010 and have baseline data going back to 2008.

The process items are important because they factor into determining appropriateness. For example, the previous conversation, if the patient's prep is inadequate, it is inappropriate to wait 10 years to repeat that procedure.

So the process items are an assessment of a standardized ASA assessment of medical risk, standardized assessment of the bowel preparation, a complete examination to the cecum, with documentation, photo-documentation of that, that if the

patient had a polyp that was removed, all of the necessary information -- there are five elements -- are completed when that polyp is sent to pathology.

The withdrawal time is recorded. The patient does not suffer serious complications of either perforation, death, admission to hospital, or bleeding requiring transfusion, and the patient was told to come back in an interval that is appropriate based on their pathology findings. So that last item will have precisely the same characteristics as the first one, although the last item, the indications for future, is future-looking.

The measure is -- I think our documents that we submitted are quite thorough in terms of the evidence for each of these items. It's grounded on the guidelines, the guidelines being those that are themselves evidence-based.

Other comments I would make would

be that the -- in our experience, based on the performance of this measure, the largest gaps are in appropriate indication at baseline. In aggregate in our data set, we had one in five patients being screened too soon. There was an occasional patient at baseline in the follow-up item that was being told to come back at too long an interval, but that was really a rarity by comparison by being told to come back too early.

So the first item and the last item were significant performance gaps in the range of 1 in 5, and the other item that had a significant performance gap was the completeness of the information that the pathologist got related to the polyp. That was a gap about 1 out of 4 times.

There's quite a bit of variability in performance across physicians. At baseline on the composite score, meaning that everything was meeting standards, the range was zero to 80, so there was -- the lowest

performer met everything in none of his patients, and the highest one met everything in 80 percent of their patients.

What we have seen over time is that, because of the structure and the precision of the measure, the history-taking and documentation has improved substantially, practice patterns have changed, and the gaps on everything have narrowed.

I believe you have with you presentation material on this, and there's trend data that's quarterly over the past two-year cycle, as well as individual proceduralist performance information.

What else would I say? I would say that we have a lot of field experience with this measure, and our experience is that the current electronic medical records are weak in the area of family and personal history, and so calculating appropriateness, both front-end and back-end, are more challenging than they should be. And so to do

eSpecifications at this point, the GI records are really not up to it. Neither are the registries.

CO-CHAIR BASKIN: Gail?

DR. AMUNDSON: Yes?

CO-CHAIR BASKIN: Any last statement? I realize it's a complex measure, but we try and keep the comments short, please.

DR. AMUNDSON: I thought I was being pretty brief.

CO-CHAIR BASKIN: You were doing great, but we're getting a little over.

DR. AMUNDSON: Okay. And then the other item on the electronic record is that the nursing and the physician components don't link. So we've been working on that.

And I think I will stop there and open it up to your questions and comments.

CO-CHAIR BASKIN: Any particular question for Gail, before we start to proceed and talk about how we're going to do this?

Zahid, do you have a question?

MEMBER BUTT: Yes. I have a question. In the data that you have submitted, you have shown stratified data, but you have chosen not to stratify the measure. Any particular reason?

DR. AMUNDSON: Not to stratify the measure?

MEMBER BUTT: Yes.

DR. AMUNDSON: Meaning?

MEMBER BUTT: Meaning to break out the rates for each individual element, like you show in your results.

DR. AMUNDSON: Well, we do have the rates for each individual element.

MEMBER BUTT: Wouldn't that fall under sort of stratification?

DR. AMUNDSON: All of the results are reported online, and they are reported by individual physicians. And so it is -- one of the screenshots in the presentation should demonstrate the drop-down list that shows that

all of the rates are also reported by individual element.

So the composite is available, but all of the elements are available for analysis as well.

MEMBER BUTT: Okay. Thank you.

CO-CHAIR BASKIN: Now, before we review this measure, this is a little bit different than the measures we have been reviewing, so I think maybe Taroon would like to make some comments. This is a composite measure, obviously, that has nine different measures within it, and does require a different level of analysis than just at the composite level.

Taroon, do you want to make some comments?

MR. AMIN: I'll actually turn it over to Karen, just to give us a brief introduction, a little bit on how we should be thinking about evaluating a composite in reference to each of the individual

components.

DR. PACE: So I just wanted to kind of orient you that NQF has done some work in the past on composite measure evaluation framework, and actually this fall we'll be revisiting that. But I just wanted to mention that we consider a composite a measure that combines multiple components, either individual measures or, as you can see here, multiple components, that result in a single score.

So, that is our definition of a composite. However, one of the considerations that we have, which kind of gets discussed later on, is the ability to decompose the composite, to look at the individual elements. So the fact that this is constructed as a single score, to result in a single score, that's our very definition of a composite.

There are definitely different types of composites. As Gail mentioned, this is an all-or-none, meaning that this

information is aggregated at the patient level. So it's looking at each patient, whether all components were met. There are other types of composites where you're actually using individual performance measure scores and aggregating them in some way, but this is an all-or-none.

And there has been a drive at NQF for calling for more composite measures, for a couple reasons. One is that it's considered a higher bar, these types of all-or-none. For example, your discussion yesterday about a measure of a patient being assessed, and then a separate measure for a patient being counseled. Well, it doesn't make a lot of sense that a patient isn't assessed and counseled.

And so these all-or-none composites are trying to get at that element, and also there's some thinking and analysis that shows that you get a stronger, more reliable, quality signal when you have more

data that goes into a performance measure.

But having said that, you need first of all to look at the overall composite, but you also need to be looking at the components, and whether those make sense in terms of the criteria that you're looking at in this aspect of importance to measure and report.

So again, the impact, performance gap and evidence. And we did ask the developers to -- we'll have to think about this in the future, but we did ask them to try to address the evidence in separate forms. So I know that created more paper, but to try to break those out for your review.

So I'll stop there and see if you have any questions about composites in general.

MEMBER BUTT: So in terms of a composite score, if you will, my sort of understanding of that was exactly the way you were describing, that you would score

individual components with perhaps even different weighting, depending on how important it was to the overall score. With this --

CO-CHAIR BASKIN: That's not necessary. That's in the eyes of whoever's doing the scoring.

MEMBER BUTT: I understand.

CO-CHAIR BASKIN: It doesn't have to work that way.

MEMBER BUTT: I'm saying that this, to me, looks more like a percentage of patients that received quote unquote "perfect scores," perfect care, if you will. So is the term scoring accurate in this context, would be my first question. Because I understand scoring to be somewhat different.

CO-CHAIR BASKIN: Scoring can be multiple methodologies.

MEMBER BUTT: So for instance, what this conveys -- it weights everything equally, right? So for instance, someone

could have a 90 percent score, and out of 100 colonoscopies they could have had 10 perforations, and the other person with a 90 percent score, the only 10 cases that could have fallen out were that they didn't take a picture of the cecum. They'd all be looked at in the public the same, according to this methodologies.

CO-CHAIR BASKIN: And there's pros and cons to different methodologies of scoring. But as this one is presented --

MEMBER BUTT: It is presented as it is.

CO-CHAIR BASKIN: It is presented as all or none. You either --

MEMBER BUTT: So it is accurate to call this a score.

CO-CHAIR BASKIN: Yes, it is.

MEMBER BUTT: Okay. That was my question.

DR. AMUNDSON: But the description of this being "perfect care" is accurate.

CO-CHAIR BASKIN: John, I think you were up first.

MEMBER MORTON: My only question is to Karen. I know there's been more emphasis in NQF about these composite measures. Can you guide us, are there other measures that have been approved, just out of curiosity, what those were like?

DR. PACE: We have some composites that are all-or-none. So, for example, optimal diabetes care, or optimal cardiovascular care, that has not this many components but multiple components, maybe five or six.

And then we have some composites -- I guess the ones that come first to mind are the AHRQ composites. They'll have a mortality composite based on procedures. But in that case, their composites are taking the individual mortality scores and combining them in some way. And I can't tell you offhand, but it could be an average of those various

mortality scores, it could be a weighted average.

So there's a variety of ways, and that's one of the things that we'll be addressing in our project coming up this fall, is looking at those different types of composites more closely, and what implications there are for those. But we have multiple examples.

CO-CHAIR BASKIN: Judith?

MEMBER TOBIN: So composites can be challenging, and I would maybe just ask the group -- I'm looking at this composite in number six, "All essential polyp information recorded." And if the group is supposed to evaluate each component of that, is that adequate? Or is that standardized enough that everyone feels they'd come to the same conclusion?

CO-CHAIR BASKIN: Well, that's going to be part of our discussion as -- I think the way we'll work through this is

probably a component at a time.

MEMBER TOBIN: Okay.

CO-CHAIR BASKIN: Because we may come to the conclusion that overall this is fine, or we may come to the conclusion that six of the components meet our criteria and three of them don't meet our criteria, and I think that's probably going to be the simplest way to work through this.

MEMBER TOBIN: Okay.

CO-CHAIR BASKIN: Are there comments about the procedure on this and composites in general?

Go ahead, Jenifer.

MEMBER LIGHTDALE: Just a question, and you may have mentioned it. So this is a process metric, the way that they're describing it. But it looks to me like it combines a process and an outcome.

DR. PACE: Right. Unfortunately, we didn't have a category for them to select something differently. But it is a mostly

process, but there is one component that's an outcome.

And actually, we do have another example of a composite that includes process and outcome, and that's the STS cardiovascular surgery composite measure. But I think that's a question that we're going to address in the future, is "When should you combine process and outcomes?" But we don't have any reason not to at this point.

And I'll just mention one other thing. And again, I know this will be hard, just as we've had some difficulty with the more single concept measures, which is that the next stage is where we would see some analysis related to these components as well. I mean, obviously you've seen the performance that they've reported on the various components, so that's useful as well.

CO-CHAIR BASKIN: Let's try and go in some order. So I think Philip, you had a comment next?

MEMBER SCHOENFELD: I was just going to suggest that I think there's a good chance a lot of these issues might be addressed once we begin to review each step.

CO-CHAIR BASKIN: Yes, that's why I'm trying to keep this to just -- because this is a different way to evaluate, so that we understand what we're evaluating and how we're going to do it, but not get into any of the details of the measure. But if there's still a question about that, please, Zahid.

MEMBER BUTT: Sure. I just have, again, one procedural/informational question. So according to -- definitionally, is there a definition for an index, if you will, or does it qualify for that definition?

DR. PACE: And that's part of the confusion out there, is that composites are referred to in multiple ways. And sometimes index is often used to refer to a composite. And as I said, our definition is a measure that has multiple components that end up in a

single score.

MEMBER BUTT: Right. So this type of composite could qualify as an index.

DR. PACE: Yes.

CO-CHAIR BASKIN: Yes, but index is a term used outside of NQF, in terms of -- you can call it what you want.

MEMBER BUTT: I was just asking if there was a specific definition that they had in that context.

CO-CHAIR BASKIN: Jenifer, did you still have a comment, or was that just up? Okay. Any other comment before we start the review process? I'm sorry, go ahead.

MEMBER BORDEIANOU: I just wanted to say that this comment about what's an index and what's a score is important. Because we think in general of indexes as scores, validative measures where one thing is not measured twice and weighted twice. And in this score, there is cecal photo and complete examination. Those are two of the same thing,

really.

So it is important, as we look at that, not only to look at it individually, but also, in the end, as a whole, so that some things are not double-counted and some things are under-counted.

CO-CHAIR BASKIN: Yes, we can certainly review that when we review the measures and make comments. But at the end, you may also make a comment to the developer that you think the name is not appropriate for what it is, or it is appropriate for what it is.

But I think that's where the term index will come in. It's really not our review in terms of the parts of the measure and the composite of the measure. But if you think the name is inappropriate, it doesn't reflect the actual measure, then that's -- we're all welcome to make those comments.

So with that, I think we're going to get started. This will be -- well, we have

some time. Fasten your seatbelts.

(Laughter.)

CO-CHAIR BASKIN: And who won this one? Phil. I hope you've done your homework. So I can only think that the way to do this is to try and break it down into individual components one at a time, and then at some point talk about how the components interact with each other, and whether they're appropriate to interact with each other, and then, in a sense, talk about it as a composite for a yea or nay. But I think we really need to do a yea or nay and vote each individual component.

MEMBER SCHOENFELD: I guess I would actually suggest that, with respect to impact, that it can be done -- we can discuss whether or not the impact of a composite score or composite index, depending on terminology, for high quality colonoscopy for colon cancer screening and colon polyp surveillance can be done without reviewing individual measures.

Then, when we go to the evidence to see if the specific measures included in this specific composite index, actually, we have evidence to show that that represents a high quality colonoscopy, that we would have to go through step by step. But it's up to you.

CO-CHAIR BASKIN: I agree somewhat, but I think what we may find is that it's possible that individual measures within the composite may not individually have any documented impact, high impact, which would make that particular portion of the measure --

DR. PACE: But I think what you're referring to is more along the evidence line. I think with impact, we're really talking about measuring colonoscopy quality in patients at risk for colon cancer, and so I think it really --

CO-CHAIR BASKIN: Good point. I probably am mixing it up.

DR. PACE: I know, it's hard.

CO-CHAIR BASKIN: So I think that

that's fair, and unless somebody has an objection to that, I think we'll discuss the impact of the composite in general, and the issue in general.

CO-CHAIR SAIGAL: That's great. Is this a competing measure? I mean, a renewal measure or a new measure.

CO-CHAIR BASKIN: New.

CO-CHAIR SAIGAL: Okay, thanks.

CO-CHAIR BASKIN: With that, Philip, impact of this particular composite?

MEMBER SCHOENFELD: Okay. With respect to impact, obviously colonoscopy for colon cancer screening and for colon polyp surveillance is performed in tens of thousands of people each year.

And that there is a growing body of data that demonstrates that missed or interval cancers do occur, meaning by definition that a patient is diagnosed with a colon cancer sooner than repeat colonoscopy would be done based on guidelines. For

example, somebody gets diagnosed for colon cancer within two years of when they had a normal screening colonoscopy.

And database studies have demonstrated that different criteria, such as not reaching the cecum, doing a complete exam in that way, are associated with an increased likelihood of having interval or missed cancers. I would note, this is not specifically part of this application.

So with respect to the idea that this is a national health goal, to do good colorectal cancer screening with colonoscopy, and that it's done in a huge group of people each year, with significant resource use, and that, to summarize, the overall goal of this composite index is to allow the consumer or the payer to say, as a yes or no question, does this endoscopist do a high-quality colonoscopy, my comment would be "Yes, that seems to be a high-impact goal."

And in fact, among the

professional GI organizations, there is a big effort to determine criteria to define high-quality colonoscopy for colon cancer screening.

CO-CHAIR BASKIN: Comments? Johannes?

MEMBER KOCH: So I think that the real outcome that we want to identify is that people don't miss cancers, right? The surrogate marker is that we don't miss polyps. Neither of those two are really -- your adenoma detection rate and your missed cancer rate are not part of this composite. So yes, I think measuring quality colonoscopy and identifying good colonoscopists versus not is really, really important, and there's a high impact of that. I'm not sure that this measure actually incorporates the two most important things, which are missed cancers and adenoma detection rate.

CO-CHAIR BASKIN: John?

DR. AMUNDSON: Can I comment on

that?

CO-CHAIR BASKIN: Actually not, Gail. I'm sorry. It's not the usual process. But if we have a specific question for you, if you're on the line, we'll certainly ask it.

MEMBER MORTON: My question is mainly about impact. Is there about 500,000 colonoscopies done a year? It's an extremely common procedure.

MEMBER SCHOENFELD: The difficulty -- and I was reviewing this again last night -- sometimes it's a little bit tough to piece out the proportion that are done for colon cancer screening and colon polyp surveillance. Even though you would think, based on reporting, it shouldn't be that tough to do, but it's actually a little bit tough to get the right estimate.

CO-CHAIR BASKIN: Other comments regarding impact? You know, I think Johannes's point is a good one, but I think there's different parts of this measure. So

yes, some parts of the measure are having performed a complete colonoscopy under good circumstances, like good prep and all that, and that doesn't get at a lot of the things that you're talking about for quality, but there are parts of this that have to do with appropriate surveillance time, similar to some of these same issues with the previous ones.

So I think there are certainly parts of this that I would say are the impactful ones that we've been discussing over the past hour or so. Parts of it may or may not be. But I think, overall, because parts of it are, my feeling is that this is high impact.

MEMBER KOCH: Just as a clarification, so the question is, is it high impact to have a composite measure of quality, or is it this composite measure of quality?

DR. PACE: Remember, because this isn't different than yesterday or the other measures, you're talking about whether the

area of colonoscopy for cancer screening is a high-impact area. You'll be getting at the specifics when you look at the evidence for what's included.

But just as you all talked about, having a measure of assessment isn't going to be as impactful as if you had a measure of actually treating something. It's in the same realm as that, right? Right now we're just talking about the broad area, that this should have a performance measure.

CO-CHAIR BASKIN: Right, we're getting into the next part of the discussion too early. So Zahid first, and then Robert?

MEMBER BUTT: And again, this might be a next-section comment, but I was just reading the summary of evidence of high-impact, and all it refers to is underuse or overuse of colonoscopy, although they do cite several studies, so I didn't obviously -- what's your take on that?

MEMBER SCHOENFELD: That's why I

mentioned, I think, at one point during my brief discussion there, that it actually wasn't part of the application. Colonoscopy is done very frequently for colon cancer screening and colon polyp surveillance. Reducing colorectal cancer, which is the second most common cancer in the United States, justifies the fact -- this is a major national health goal, just reading off the definition of national goals and priorities. It's high impact because a lot of resources are used, and in terms of consequences, if you're not performing a high-quality colonoscopy, people get interval colon cancers.

So again, just to reinforce what Karen said, I'm not commenting about the specific components of this colonoscopy quality index. I'm merely talking about, is there a rationale to say "A colonoscopy quality index that allows you to tell if somebody's doing a high-quality colonoscopy to

minimize or prevent colon cancer," is that potentially a high-impact quality indicator? And my interpretation is yes.

CO-CHAIR BASKIN: Thank you. And Robert?

MEMBER ELLIS: I think it's important, as kind of the consumer guy on this group, to point out that these composite measures, at the consumer level, are your best shot at getting something even looked at. It's very difficult to get consumers engaged in looking at these scores, indexes, however you want to characterize them.

Composites are your best shot. There's a lot of literature. Our own studies of the stuff we deliver, composites are usually easier for them to wrap their brains around. They usually answer simple questions. They're interested in "Did things get done the right way, and did all the things that are supposed to get done, get done?" That they can understand. So I think these composites

are really important to really speak to consumers.

CO-CHAIR BASKIN: Okay. So, certainly an impact on the consumers. I think we've spoken enough about this that I think we can come to a vote on impact. Once again, it's going to be a hand vote. We have a 1, 2, 3 or 4, so those voting 1, high impact?

(Show of hands.)

CO-CHAIR BASKIN: Looks unanimous. So that's 15 high impact, and obviously zero for the other categories.

Okay. So now it gets difficult, right? Philip, I'm hoping that you organized your thoughts one portion of the measure at a time, but let's go that way.

And I really do think, for the ease of -- I mean, otherwise, we're going to get off on tangents like crazy. If we really try and stick to each component, because there will be an opportunity at the end for us to talk about the components and how they

interact with each other, but if we talk about it with each component before we've talked about the other component, I think we're going to drive ourselves crazy.

So let's try and -- just because we voted yea or nay for nine components doesn't mean yea or nay for the composite, and there will be plenty of opportunity to discuss what should, may or may not, fall out, and what works with what, and what interacts with what in the right way to get where we want to be. So let's try and minimize this, leave this to just the component itself, when we're talking about it. Thank you.

MEMBER SCHOENFELD: I may be foolishly hopeful that this will be easier than you anticipate, Andy.

I'll start my comments by re-emphasizing what Mr. Ellis said. Having a composite index to allow -- that is a yes or no endpoint, to say that an endoscopist has performed a high-quality colonoscopy, is a

crucially important quality indicator, that we want consumers and payers to be able to say "Does an endoscopist perform a high-quality colonoscopy 90 percent of the time," and let the consumer have that, and then for the payers, also be able to look at these different subcomponents.

And I really commend Quality Quest for Health of Illinois for putting forth this packet. Having said that, the components of this quality index, in my opinion, do not have the evidence support to justify multiple components of this quality index. So in other words -- and I'll go through each indication very briefly, and then we can begin, if you'd like, to do the vote on each indication. To quote an old phrase, it's just not ready for prime time yet.

So having said that, let me begin to go through each one very briefly. Certainly the first thing is that they say that you should document an appropriate

indication for colonoscopy. We had similar discussions about this yesterday. I mean, this is a procedure. Maybe not quite like putting somebody to sleep to do a surgery for stress urinary incontinence, but good medical practice is, of course you have to write down an indication. And so I'll just keep that one brief.

You also have to do a standardized medical risk assessment and document that prior to doing a procedure. And that actually is part of the PCPI measure set that's already been put forward as of 2008.

The next thing, though -- so, this is where we begin. Those two, to me, are almost pro forma. But having said that, the next one becomes a little bit more difficult, or the next point about standardized medical risk becomes a little bit more difficult.

As part of your standardized medical risk assessment, you do a cardiac risk assessment. The problem is that in many

practices right now, anesthetists do your cardiac risk assessment, because actually nurse anesthetists and anesthesiologists actually provide the anesthesia. So the medical risk assessment isn't really done purely by the endoscopists. It's done by two people.

So, how do you actually operationalize that part of your assessment? I'm sure that Quality Quest for Health Care, within their system in Illinois, has a mechanism for doing that. Whether or not that could be implemented nationally, I think, may be a different issue.

The really big issues, though, come once we start talking about assessment of bowel preparation, standardized assessment of bowel preparation, complete examination, cecal photo taken, all polyp information recorded, withdrawal time recorded.

What they're trying to get at is, does an endoscopist actually get all the way

around to the cecum, state that all the stool has been cleansed out, or most of the stool has been cleansed out, so that they can adequately identify polyps, document how many polyps they found, and also document how long they spent pulling the scope out?

Ultimately, those are trying to get at the issue of whether or not you're doing a colonoscopy that's going to minimize or prevent somebody from getting colon cancer in the future. The problem is the way they have the evidence here, that documenting these factors is going to lead to the outcome we're hoping to get, which is people aren't going to get colon cancers in the next couple of years after your colonoscopy. The way they're outlined here is not going to achieve that outcome. So let me begin to specifically go through this.

Should a composite index in the future assess bowel preparation? Absolutely. If somebody routinely -- this was already

mentioned by Johannes -- documents that they get a bad bowel prep, and thus justifies -- because that would be a preparation for repeating the colonoscopy again a month later, or within a year -- that's actually an indicator that, within their practice, they're not doing a good job of colonoscopy. You know, three quarters or more of your patients should really have an excellent bowel prep, and if virtually all your patients have a poor bowel prep, then you're not doing your practice properly. So that part, I go along with.

Complete examination means you get all the way to the cecum, and they try to subsume that with the idea of taking a photograph of the cecum. There's not necessarily great evidence to say that that definitely support the idea you get to the cecum, but we generally accept that as a standard of practice, that if you get to the cecum and take a photo of it, that proves --

or certainly the appendiceal orifice, as well as the ileocecal valve -- that that proves that you got there.

So again, I can go along with that part. But the ultimate thing, besides getting all the way to the cecum, that Johannes mentioned, is that then we want to show that you're adequately identifying adenomas.

Now, what they talk about here to show that you have an adequate adenoma detection rate is how long you spent pulling the scope out, and describing the size and shape and location of any polyps that you found.

That does not tell you whether or not you're finding adenomas. Withdrawal time, how long you spent pulling it out, impacts adenoma detection rate. If I actually quantify how often somebody finds adenomas -- and the current guidelines recommend that it be found in at least 15 percent of women and 25 percent of men -- if my withdrawal time is

4 or 5 minutes, and my adenoma detection rate is 5 percent, well, okay. That's something I can identify to work on.

But lots of people have withdrawal times of 10 minutes -- and by the way, the cutoff is felt to be seven minutes. Lots of people have withdrawal times of 10 minutes, and they still only find adenomas in six or seven percent of people.

CO-CHAIR SAIGAL: Phil, can I ask a question real quick?

MEMBER SCHOENFELD: Sure.

CO-CHAIR SAIGAL: This is really helpful. Can you let us know what level of evidence is supporting what you're saying? Is it consensus, is it in the document? That would be really helpful to understand.

MEMBER SCHOENFELD: Okay. So just to then briefly go back, for appropriate indication for colonoscopy, documenting that, I would say that is consensus opinion.

For standardized medical risk

assessment, meaning that prior to performing a procedure you do an appropriate cardiac/pulmonary risk assessment, that's standard of care consensus opinion, and I would merely point out that documenting that is problematic to the extent that two different providers are doing that assessment.

With respect to assessment of bowel preparation, we actually have multiple endoscopic database studies and randomized controlled trials that demonstrate that the quality of your bowel preparation impacts your adenoma detection rate.

So in other words, if I get an excellent bowel prep compared to what we call a fair bowel prep, meaning it's not so horrible I have to immediately repeat it, but I'm only visualizing about 80 percent of the mucosa instead of 100 percent, that when I have an excellent bowel prep, my adenoma detection rate increases by two- to three-fold.

CO-CHAIR BASKIN: Is that evidence submitted?

MEMBER SCHOENFELD: No, that's not part of their application.

CO-CHAIR BASKIN: Okay.

MEMBER SCHOENFELD: I'm expanding here on what I know to be the data. So you have RCT data on that, as well as database data to show that quality of bowel preparation is associated in the database data with higher adenoma detection rates.

With respect to complete examination, we have database data to demonstrate that. Failure to reach the cecum is associated with a higher risk of having interval cancers.

CO-CHAIR SAIGAL: Is that in the document?

MEMBER SCHOENFELD: No, that's what I know to be the case. So that's why we feel it's important as a quality indicator to be able to document that you reached the

cecum.

Having all essential polyp information recorded, having your withdrawal time recorded. Again, the idea there is, does that equate to finding adenomas? For withdrawal time, we have endoscopic database studies that demonstrate that, if your withdrawal time is greater than seven minutes, your adenoma detection rate is higher than if your withdrawal time is less than seven minutes.

However, what we also know from endoscopic database studies is that plenty of individuals who have withdrawal times of greater than seven minutes are still poor performers in terms of adenoma detection rate. Recording withdrawal time is helpful, because if somebody's a poor performer and they have a very low withdrawal time that's something to work on. But it doesn't encompass the bottom line, which is we need to know if people find adenomas.

And by the way, on that one, we again have good endoscopic database data to show people with higher adenoma detection rates have fewer interval cancers. People with lower adenoma detection rates, they're more likely to have patients who have interval cancers. So the key there would actually be to record an adenoma detection rate.

CO-CHAIR BASKIN: But if you're recording all essential information about polyps being found --

MEMBER SCHOENFELD: But you're not. The key piece of information that's not included -- here you're saying you describe the size of the polyp, the shape of the polyp, the location of the polyp, and how you removed it. What's not there is the actual histology of the polyp.

CO-CHAIR BASKIN: Which you don't know at the time of the colonoscopy.

MEMBER SCHOENFELD: Which you don't know at the time of the colonoscopy. So

that an appropriate quality indicator for adenoma detection rate is going to require people to be able to document in their database, in their registry, the results of the histology of the polyps they removed.

Free of serious complication, this would be consensus, standard of practice, that you document whether or not you perforated the colon at the time you did the colonoscopy. I would simply note that, in a normal screening colonoscopy, the likelihood of getting a perforation, based on meta-analysis, is 1 in 3,000 or less.

What's really the key with documenting complication is being able to follow the patient out for 14 to 30 days. The vast majority of complications, bleeding after you've taken out a polyp, they occur 24 hours or more after the person's had the colonoscopy.

So my point about that is, and this is consensus, documenting whether or not

you've had a serious complication at the time of the colonoscopy is definitely important to do, but it's not actually getting at a true measurement of complications, which would require having follow-up with the patient in 14 to 30 days.

CO-CHAIR BASKIN: And would you also say to that, similar to what you've said about appropriate indications, standard assessment, that any procedure that is performed, if you know of a complication at the time of procedure, it's standard to document that?

MEMBER SCHOENFELD: And so that part is definitely consensus, that it's standard of care to report it at the time of the procedure. My point about this would be that having a complication at the time of the actual colonoscopy is exceedingly rare, and complications from colonoscopy, the vast majority occur in the 14 days after the colonoscopy has been performed.

Again, just to provide one example, if I take off a big polyp, it's rare that it bleeds significantly right at the time of the colonoscopy that I can't control. Most of the time, it doesn't bleed at all. But if I'm going to get a post-polypectomy bleed that leads to hospitalization, that usually occurs 3 to 14 days after the colonoscopy. You can't document that based on this kind of a quality indicator.

CO-CHAIR SAIGAL: So a few things are being discussed here. It sounds like, in general, most of these don't meet the NQF standard for evidence. A lot of consensus stuff.

MEMBER SCHOENFELD: Correct.

CO-CHAIR SAIGAL: For the ones where there is evidence that you're aware of, it's not in the document. Like, for example, database or observational studies about withdrawal time, even though they have withdrawal time recorded, not even the cutoffs

that you mentioned.

You also brought up issues that go to Johannes's point about importance, really. Because it sounds like what you're saying is that, unless you know the histology of the lesions being removed, you can't make an inference about the quality of the colonoscopy. So that has to do more with the importance of the measure, which we already voted on, but it's sort of in that ballpark, still.

And then you also -- in terms of importance, you won't be able to measure complications at the same sitting, basically. So that also goes to importance, I think, because you're saying there's not enough of an opportunity to measure outcomes important to patients.

So, I don't know if that was clear to the group when we voted on importance, but those are two things that --

DR. PACE: You voted on impact.

CO-CHAIR SAIGAL: Impact, right.

DR. PACE: All of this is related to importance. All three of these.

CO-CHAIR SAIGAL: I'm sorry, I meant impact. But those are impact. So basically, what you're saying is, if you don't have the histology and if you don't have the ability to follow up for 30 days, that this might be a low-impact measure, is what you both are advising?

MEMBER SCHOENFELD: Well, I think I would say that, in terms of the evidence, evidence that documenting serious complication at the time of colonoscopy demonstrates that somebody has a high-quality colonoscopy has a very low impact.

Again, the general theme that we voted on: is it really important to have a colonoscopy quality indicator, a composite? Absolutely high-impact. When we look at the specific components here, is there good evidence to say that documenting serious

complications at the time of colonoscopy is a good representation of complication rates from colonoscopy?

What I'm saying is that the evidence doesn't support that. You do need to document it at the time of the procedure, but again, perforation, 1 in 3,000? That doesn't really get at measuring frequency of complications from colonoscopy.

CO-CHAIR SAIGAL: My understanding of voting on impact, then, was that we vote on the impact of the measure that's in front of us, not the idea.

DR. PACE: Let me. Impact is about the general area, as Phil's just been saying, of having a quality measure about colonoscopy quality. And what we're getting at through the other criteria of performance gap and evidence is whether there's evidence to support that particular component, what's being measured, or there's a performance gap. And those three things together combined to

designate our importance to measure and report category.

CO-CHAIR SAIGAL: Okay. Thank you.

DR. PACE: So impact is much more general.

CO-CHAIR SAIGAL: All right. Thanks.

MEMBER SCHOENFELD: Karen may comment on this more. I think the confusing part is, this is the first composite index we're looking at, as opposed to previous ones we have looked at that just look at one specific question, where there's much better correlation between impact and the evidence associated with that.

CO-CHAIR BASKIN: I know there are other comments, but you only had one more component to talk about before the comments, so maybe if you can make that, and then the others can start to comment? Because you got that far, I hate to break it up.

MEMBER SCHOENFELD: Appropriate follow-up recommendation is something that CMS has already included in PQRS. That's definitely very appropriate, and we already talked about appropriate follow-up recommendations, actually, within the last two indicators.

The reason I went through all these is that, in summary, it appears to me, my recommendation as the lead discussant is that when we get to the question for evidence, about whether or not the evidence supports this specific colonoscopy quality index, when you look at all these factors, the answer's going to be no. So that's why I kind of did it in this way.

Having said that, I mean, everybody's going to vote the way they want to vote, make comments the way they want to vote. But I think having a long discussion about each individual one, in my opinion, may not be necessary.

So I'll go ahead and stop there. I've talked enough here.

CO-CHAIR BASKIN: So there were a couple comments about that overall assessment. And thank you, Philip, for going through that in a stepwise fashion. But John, you had wanted to make a comment first.

MEMBER MORTON: I guess my only comment is, procedurally, how we're going to go about this. Are we going to be all or none in this measure, or are we going to look at them step by step? Because there's a lot to kind of go through here.

CO-CHAIR BASKIN: I do think, at the end of the day, we have to do all or none in terms of a final decision as to whether to move this forward. But I do think, in terms of feedback to this developer, there may be feedback about individual components which could strengthen this measure to come back at another time if the entire measure fails, which obviously is still going to be up for

vote, but there's obviously some concerns about many of the components. So I do think we're going to have to get into the individual components and give some feedback to that.

Zahid, you wanted to make a comment?

DR. AMUNDSON: Andy, this is Gail. I need to make some corrections to the -- I just feel there's a need to make some corrections.

CO-CHAIR BASKIN: Gail, I was going to allow us to ask you some questions, but that would be helpful. If you just keep them brief, a few comments would be helpful to us. Thank you.

DR. AMUNDSON: Right. So the complications, it's not documenting complication or lack thereof. It's the patient -- the procedure fails quality if there is a serious complication within 24 hours. And we have the data on that, and it's not 1 in 3,000. It is low, but it's not a 1

in 3,000.

I would say the recording the time, the element is not a specification of what the time should be. It's rather that the procedure records the time. And most of the state of the art procedures these days have an electronic timestamp, because the GI community has agreed that that's important.

The ASA is the American Society of Anesthesiologists, and some of those shouldn't be getting screened.

The information on the polyp is so the pathologist has the information they need to make an accurate pathology interpretation. The challenge is that the pathologist doesn't know whether it was a complete or a partial polyp removal. This is a very big and important issue. And I'd say to suggest that there's not evidence to support each one of these is not accurate.

CO-CHAIR BASKIN: Okay. Thank you, Gail.

DR. AMUNDSON: One other comment is the adenoma detection rate is not appropriate to be -- and the pathology is in this measure because that's how the follow-up indications -- so the pathology is known before this measure is completed. We do have adenoma detection rates, but those are rates across 100 patients. They're not for any one individual patient, you know? That just doesn't work that way. You can't have a population rate at an individual level. This is an individual, person-level, did this patient get a high-quality procedure, yes or no?

CO-CHAIR BASKIN: I appreciate that. So it's really not a measure of a population of folks to see whether an endoscopist is -- well, I guess you put it --

DR. AMUNDSON: It is both, Andy, a population and an individual. But an adenoma detection rate is not an individual, person-level measure.

CO-CHAIR BASKIN: All right. Zahid, you wanted to make a comment? Thank you, Gail.

MEMBER BUTT: Yes, I think Gail made a couple of those points that I was going to make, but I think in general my comment is that, really, the intent of this measure is to understand whether a quality colonoscopy was done or not. We already said in the impact that it is important for a quality colonoscopy to be done for all the various reasons, but the question here is, do these components represent a quality colonoscopy, and is there evidence that is presented to support that? So I think, really, that's what we need to focus on.

And in terms of the evidence itself, it was somewhat difficult the way it's sort of presented. And as I said, maybe it's an issue of not that there isn't evidence, but that it isn't -- at least when I read it, it's sort of the same thing is repeated over and

over again, just those two or three studies, and the guidelines are repeated over and over again, not specifically referencing each section, although each section is broken down. It doesn't really address each section, that this is the evidence for this section.

And I think from our sort of collective experience we can say that pretty much all of these things are fairly standard type of quality things that are measured when you're just trying to understand whether a quality colonoscopy was done or not. So I think we should probably really focus on that aspect, as to whether there is evidence presented for each one of those. And for the ones where there isn't evidence presented, is there other evidence out there?

And we'll sort of get into that, again, tricky issue of, how do you deal with it when a few of them might be a 3, and three might be a 1, or maybe one might be a 1 and four are a 2, and so that sort of is a

challenge that we'd have to address.

But I think really the focus -- the point I was trying to make is that the focus should be whether these represent, in aggregate, a quality colonoscopy. So whether it's adenoma versus hyperplasia doesn't matter, really, in that context.

CO-CHAIR BASKIN: Johannes?

MEMBER KOCH: I think that adenoma detection rate is really a crucial metric here. And part of what's happened is, in the GI field over the last 10 years, we started by looking for surrogate markers for adenoma detection rate, and have all really agreed -- and the data, large population-based studies have proven that adenoma detection rate is the single most relevant predictor of patient outcomes.

So we looked at withdrawal times and adequate polyp preps, and all the other reasons to do a colonoscopy or to look for quality, and they are surrogate markers of

adenoma detection rate. So we have the best marker, which is "What percent of patients do you find an adenoma on?" Beyond that, we know that "What percent of patients do you find an advanced adenoma on?" These are all known metrics within 72 hours of having a colonoscopy done, so beyond two weeks you can find out whether or not the procedure identified a high-risk lesion or not.

And really, what we're trying to do is find out if you're a good driver. Having a car accident is missing cancer. Getting speeding is you have an adenoma detection rate that's too low. And we're using metrics like "Do you rotate your tires?" and "Do you have a clean car?" to decide whether you're a good driver. We have the metrics, which is "How many cancers do you miss?" and "What's your adenoma detection rate?"

And really, for a patient, do I care whether they said they took 10 minutes to

withdraw on my colonoscopy? I could say that. I could say your bowel prep was great. But I want to know, on the last 100 patients, did you actually find the same number of polyps?

And there's great variability. Advanced adenomas are the single most important, because those are the ones that, in the next five years, if you missed them, are going to be cancer. And that rate should be up to 10 percent, and there are people who have an advanced adenoma rate less than 1 percent.

Those are the pertinent findings. Yes, we know that if you don't get to the cecum you miss colon cancers. You miss them. So that's an important metric. However, that falls within the range of your entire procedure.

So I think that there's really good data to say that you should do an appropriate indication. There's really good data that if you do colonoscopy too soon,

you're not doing good colonoscopy. Similarly, if you recommend a colonoscopy too early, that's inappropriate. Those two are the measures that we've already addressed. I think for the rest of them, the data is for adenoma detection rate, not in any of these other metrics, to suggest that they, in fact, impact any patient outcome.

CO-CHAIR BASKIN: All right. Jenifer?

MEMBER LIGHTDALE: I mean, I agree with both of you, and I also really am taken with this concept that patients respond to composite metrics. This is clearly intended to be a patient-level metric, and I like your analogy to driving a car, and what makes a good driver.

I would worry that the standard of evidence that we need to hold this metric to is that what is being presented as the components of this metric are actually -- the evidence would be that that means it's a

high-quality colonoscopy for that patient.

And so getting back to this withdrawal time question -- really, the question here with withdrawal time is, is it a surrogate for adenoma detection rate? And the answer is no. I mean, the evidence will not -- well, it's a predictor. It's a predictor. But is it actually -- that's your entire point, Phil, right? Is that a very slow withdrawal does not mean you have a high adenoma detection rate. So I think that would be critical here.

CO-CHAIR BASKIN: Robert?

MEMBER ELLIS: Let me take the car analogy a little bit further. Maybe you can help me, because I'm wandering off a little towards the end. If we use the car analogy, there's like 38 steps in doing a valve job on a car, and there's a lot of shortcuts in those 38 steps. If you do all 38 steps, there's a pretty good chance you've done a good valve job.

I can then take those rings, valve seals, put them under an electron microscope, and make additional diagnoses about the problems with the car. And from that, I may end up actually providing the engine some benefit, right? Because I've found that it has probably a blown master gasket, or something like that.

That doesn't deflect from the fact that the mechanic did a good valve job, right? And I'm wondering where this measure's kind of endgame is, and the relative detection of adenomas, although obviously related to a very important outcome, is it the defining point of "Did you do the procedure in a quality way?" And I don't quite know where that line falls.

CO-CHAIR BASKIN: Karen wanted to make a comment before we moved in between.

DR. PACE: I am sorry, I have to leave for a brief conference call. But I just wanted to make a couple points for your consideration. And one is that the adenoma

detection rate is not the measure before us. It's about -- and if that's truly the better way to measure quality of colonoscopy, you need to think about the measures that you approved or recommended to move forward, which were simply recommending a 10-year return versus -- I mean, so you need to think about this and balance in terms of what's most important in terms of getting out a quality index.

So I'm not saying that all of these components are absolute, but you need to think about this measure as it's being presented, and a different measure might be preferable, but you need to talk about this measure in terms of what it's doing and, as you did yesterday and this morning, even if the evidence -- you can make exceptions for expert opinion for different components as you did for other, single measures.

So I just wanted to mention that as well.

CO-CHAIR BASKIN: All right. So, I partially lost track of when these signs went up, but I do know that Ed's came up first, before the others, and I do know that Philip's came up last. And I'll try to get the middle ones right. So Ed, go ahead.

MEMBER GILL: Thanks. This is just a quick question. If we're supposed to be evaluating these measures based on the evidence submitted, I need some help with this new process where their evidence equals logical argument.

(Laughter.)

MEMBER GILL: And they reference number two, the parachute hypothesis. So to me, I don't know what to do with that. That's not evidence.

DR. PACE: Right. So, according to how you voted yesterday, you either said "Yes," "No, it doesn't meet the criterion, there's no empirical evidence," and then you could invoke the exception, and the third no

was "No evidence submitted, but you're aware of a body of evidence that exists."

MEMBER GILL: Right. It seems to me that's where we are, and the rest of this is moot, and we just need to rely on our GI colleagues to help us tell if there is other evidence that would be helpful here.

CO-CHAIR BASKIN: All right. So I am going to go Johannes and then Zahid.

MEMBER KOCH: I think that we all feel very strongly that a composite index would be really, really a good thing. So I think to Karen's point, adenoma detection rate is the benchmark that should be incorporated. And if this had included that, the usefulness of this would be very, very different. I think there's markers here that are surrogate markers for that, and we know what it is. We know what the marker is. So I think that, just because we approved proper surveillance intervals doesn't mean that we should approve this, because it's a composite marker.

I think that the question -- and we're getting mixed up in analogies. The simple question, I think, is "Does the consumer care about a physician who misses cancers and has a very low adenoma detection rate more than they care about a physician who says that you had an adequate bowel prep and says that they saw all the polyps that you had?"

I mean, those are different weightings of that, and there's no weighting here. And in terms of evidence, really, for these affecting outcomes, there's very, very little evidence for most of these measures.

CO-CHAIR BASKIN: And I am going to break the chain here just for a second, because I know Gail's on the phone, and I'm going to make a guess that Gail thought about adenoma detection rates in creating this measure. And in fact, obviously, if you put all the information about polyps, and are just knowing what the results of those pathology

reports are before you give a recommendation for follow-up, it could have been calculated. So Gail, is there some reason why that specific measure is not part of the composite? Was there a reason for or against that?

DR. AMUNDSON: Well, I think you have to ask yourself, how would you put adenoma detection in a measure that -- adenoma detection rate is an important measure, and we have the data on adenoma detection rate. But it is a paired measure with this, because this is about a good valve job.

I love that analogy. This is about a good valve job, and I would really push back on there not being evidence for these things. There's evidence for every one in there. We were asked to do them each separately, which is why they're repeating. It wasn't probably our preference, because we think it makes it much harder to get into the evidence.

But the adenoma detection rate is

44 percent in men and 31 percent in women in this region, and what has happened is that, as the colonoscopy all-or-none composite drives up to consistent, high-level reliable procedures, our adenoma detection rate has skyrocketed.

And that's the reason why the process reliability is an important measure. But you can't put an adenoma detection rate inside a person-level measure. It is not possible to do it.

CO-CHAIR BASKIN: Thank you.

DR. AMUNDSON: And I would disagree with the comment that patients care about "Are you missing cancer," because patients don't understand an adenoma detection rate. We've tried that. They don't get it. They don't know what 44 percent means, and they don't know how to compare that to 15 percent.

CO-CHAIR BASKIN: Thank you, Gail. And I think part of our problem with this

discussion is that we're looking at this measure differently than the developer intended it to be. As I see it now, it sounds like what the measure's really saying is "Did you do all the parts that are necessary to make it a high likelihood that you'll get a better result?"

Not what the result is, not whether you found a lot of polyps or didn't find a lot, or adenomas, not what the end outcome is, but "Did you follow all the processes that are shown to be an essential part to make it a higher likelihood that the patient will have gotten the quality colonoscopy?" And that there would be a different measure set if you were looking at the population and saying "What are the health outcomes of that?"

And Gail's pointing out that yes, there were health outcomes improvement that were shown, of increasing adenomatous polyp detection based on the history of this, but

that's not what this measure was attempting to measure. That's in itself a separate measure. So we're mixing up outcome measures and a process measure, to say "Did all the elements of a colonoscopy occur to make it a higher likelihood that the colonoscopy was a quality colonoscopy?" The physical act of doing it, and all the components that you have to do to make sure that anything you did, you got the full information from it.

You know, did you actually tell the pathologist what size and what piece of the polyp is there, so there's a likelihood that you'll get a quality report back? That's really what this is saying. At least, that's how I'm viewing it.

I'm sorry to interject. So I think Zahid, and then Stuart.

MEMBER BUTT: So I think, really, the question that we have before us, it looks like from all this discussion, is that the components, 3 which is the bowel prep, 4 which

is complete exam, 5 cecal photo taken and 7, withdrawal time recorded -- the question is, are these four components -- basically, is the adenoma detection rate complementary to these, or these don't matter at all because the adenoma detection rate really replaces them?

So, we have to look at the evidence, that are these four components, which really are driving, presumably, the adenoma detection rate outcome -- is the evidence such that they are useless, or not helping the adenoma detection rate? Because that's really what the crux of this discussion is. Everybody agrees on the other components.

CO-CHAIR BASKIN: Yes, and I think that's what I was saying as well. In other words, if you're going to have a good adenoma detection rate, you have to have done a high-quality colonoscopy. And did you do a --

MEMBER BUTT: Well, what we have to determine here is, is there evidence that that's the case? Or is there evidence that

that's not the case, actually? Is there evidence contrary to the fact that if you do all these things, it doesn't matter to the adenoma detection rate?

CO-CHAIR BASKIN: Stuart, you've been waiting patiently.

MEMBER REYNOLDS: Sort of going along with that, my issue is I've got a 115 page document and there's almost no data or evidence in there. And so in an attempt to try to move things along, we're proposing that we're going to vote on each one of those things. And almost without exception, they're all going to be at best insufficient, if not nonexistent.

And then we're going to be faced with a vote with "Well, are we going to push it through anyway based on consensus?" And I think we should try to move along in that way, and I would call that we just start voting on these things. And as they come up, we're going to be faced with the decision, do people

feel strongly enough that they go forward? Because then we still have to get back to the composite thing as a whole, and we're really getting bogged down in these ideas?

CO-CHAIR BASKIN: I would hope that the comments could speak to that. I think it is a good idea that we start to move along, because we're starting to rehash here. But at the same time, if there's a comment that links to that, please make it.

CO-CHAIR SAIGAL: Can I just make one comment about this, and then I'll move us along? Stuart is right, basically. We're supposed to be recognizing if there's evidence supporting these measures. All the document says is "It's common sense that these work," and there's a joke reference for most of them. And the developer, I think, is not correct in advocating that there's a lot of evidence in the document for what they're saying.

The question is -- we'll be voting it down, I have a feeling, in terms of the

evidence being in the document. Will there be importance to measure as an exception? This is where I'm concerned, personally, based on what our colleagues are saying about this adenoma detection rate.

And I think there could be a patient-level measure that says "Does this doctor have an adenoma detection rate above a certain threshold?" A very low bar threshold, but you could learn from that measure, then, this is a doctor who does find adenomas, in general. So you could conceive of it as a patient-level measure.

That's my only comment. I think we should probably move on unless there's anything pressing.

CO-CHAIR BASKIN: Phil, if you wanted to say something as the presenter?

MEMBER SCHOENFELD: I was going to reinforce what Stuart said. I would suggest that we take a vote on the second criteria, on evidence. And I think it's going to be pretty

-- I think, I guess what the outcome will be. And then if we want to discuss further feedback to the developer, I think that would be fine.

But I would just reinforce, if I understood what Stuart said correctly, maybe it's time to just take a vote on that second question.

MR. AMIN: I will just jump in here real quick. One thing that I will just clarify -- and I know we've framed this in multiple different ways, and Karen said this, but I just want to make sure I reiterate it. What we are here to do is evaluate what's in front of us, and clearly the evidence question is asking us the quality, quantity and consistency of the evidence of this process measure and the components of the process measure that influence quality outcomes that are important for patients.

And I think that's the frame in which you have had the conversation, but

that's different than having a discussion around a different measure. So while those are interconnected, we want to make sure that what you're looking at right now is the evidence that supports this measure, that influences patient outcomes that matter. And let's keep it there, and not necessarily around new measures or a measure that this is not constructed to do.

DR. AMUNDSON: And the most important reference is --

CO-CHAIR BASKIN: Gail, no comments. I'm sorry, you need to shut off. And shutting off this conversation, I know, Johannes, you got it in there. So if you can keep it within 30 seconds, we'd like to get this to a vote.

MEMBER KOCH: Less than 30 seconds. I think the one component here is that many of these metrics lend themselves to gamesmanship, so that you can say things about it, and adenoma or other metrics, hardcore

metrics that can't be manipulated, I think would be more valuable.

MS. WILBON: So Andy, can I just suggest that we go through the list of the nine components, start with 1, make sure that the developer knows which? We say the name of the component?

CO-CHAIR BASKIN: I can't agree with that, only because part of this is actually surveillance recommendations, for which we just approved two measures that are very similar. And to say that the evidence?

CO-CHAIR SAIGAL: But that was for overuse. That was a different impact on the patient.

CO-CHAIR BASKIN: But nevertheless, it is a component of this. It's unclear to me whether it's not homogeneous with it. It doesn't make sense with the rest of the components, but nevertheless, it is there. And I don't think it is just whether a recommendation was made, but it was the

appropriate recommendation for follow-up, which is an outcome for which we've already said there is evidence.

CO-CHAIR SAIGAL: We said that there was evidence there wasn't harm to wait longer for the patient. We were avoiding complications of colonoscopy.

MEMBER SCHOENFELD: I've got to go back to what Taroon said. My impression is, we're supposed to get to a point where we say -- this is an all-or-none quality indicator based on multiple components, where we say "Yes, evidence supports this all-or-none index score, that it should be used as a quality indicator and go through to Stage 2," or "No, it doesn't."

We're not saying that some aspects of it aren't good. We're saying -- if I understand what you said correctly, we're voting on what's presented to us. Does this all-or-none colonoscopy quality index have the evidence to say "Yes, this is the right

all-or-none quality index?"

CO-CHAIR BASKIN: Not whether it's the right one, because it's the only one. So it's whether this --

MEMBER SCHOENFELD: Right. If this is the one --

CO-CHAIR BASKIN: If this is a good measure or not a good measure based on the evidence.

MEMBER SCHOENFELD: Yes, I accept that.

CO-CHAIR BASKIN: Okay. Well, then, I think we'll bring the all-or-none to a vote regarding the evidence base. At some point, we could always provide feedback to the developer about individual components.

So let's get the potential votes of what it is there. The body of evidence that's presented, that's been submitted, meets the guidance, or no, the evidence doesn't meet the guidance, or that it's insufficient, but perhaps we think there is a body of evidence

to support it.

There wasn't -- as reported to us and as we've read, there wasn't a whole lot of body of evidence about most of these components, and I think I hadn't heard anyone say that that body of evidence actually exists anywhere either. That's what I took from that voluminous conversation.

So I think we'll just go ahead and vote. It's a 1, 2 or 3. So those voting 1, yes, the body of evidence submitted meets the guidance, raise your hands any time now?

(No hands.)

CO-CHAIR BASKIN: Okay, that's a zero. 2, the evidence does not meet the guidance for quality, quantity and consistency?

MS. BOSSLEY: And that it doesn't exist.

CO-CHAIR BASKIN: Yes, and that it doesn't exist.

(Show of hands.)

CO-CHAIR BASKIN: Twelve. And then 3, no, insufficient evidence, but that that body of evidence may exist.

(Show of hands.)

CO-CHAIR BASKIN: Three. And that's 15, so that's fine. So now I think, then, we go to whether the -- is this the exception one. I was looking for the word exception, but I didn't look at the blue part.

So now we're trying to determine whether, despite the fact that the evidence was not submitted and that we think that it may not exist, whether there's expert opinion, or whether this is a standard acceptable and not a big leap for us to make to say that this is reasonable. And we'll open that up for a couple of comments. So, Phil?

MEMBER SCHOENFELD: I waited until this point to bring this up. Okay. The way this is phrased, "If there is no empirical evidence, only expert opinion, and expert opinion was systematically assessed with

agreement that the benefits to patients greatly outweigh the potential harms, is there an exceptional and compelling reason that the measure should be considered further?"

So now we're saying do we think, even though there's not evidence, is there expert opinion that this is so exceptional it should move forward? So I waited until now to say this. I've been involved in professional organizations in GI for a lot of years, and the three different GI organizations virtually never agree on anything. But you've got, in the public comment here, a letter signed by the presidents of all three organizations all saying that this colonoscopy quality index does not meet the criteria to assess the quality of colonoscopies.

So at least with respect to the three GI organizations that represent virtually all gastroenterologists in the world, because the American Gastrological Association is international, they all say the

answer to this question would be no.

CO-CHAIR BASKIN: And I'm presuming we've all read that, but for the sake of conversation are they a little bit more specific about why they're saying no?

MEMBER SCHOENFELD: They actually basically say "Assessing bowel prep, yes, and that's been submitted to CMS. Assessing a full exam by photographing the cecum, yes. Using withdrawal time instead of adenoma detection rate, no. Using serious complications," which if I understood Gail, is actually within 24 hours. Well, actually, you're not really saying "within 24 hours." You're documenting complications right at the time of colonoscopy.

DR. AMUNDSON: That's not accurate. That's not --

MEMBER SCHOENFELD: No. You're not an endoscopist, ma'am. That's not the case.

CO-CHAIR BASKIN: Whoa, slow down.

So Gail, I'm sorry, but this is not an interactive conversation. I appreciate that you're champing at the bit to say something. How this is actually measured and when the measure is actually reported as to whether you submit your data after 24 hours, and you really do get a 24 hour complication rate, that's got to do with the feasibility and the ability to report this.

But I don't think this necessarily says "You completed the colonoscopy. Is there a complication?" If it measures 24 hours worth of complications, it measures 24 hours worth of complications, assuming you're doing it correctly.

MEMBER SCHOENFELD: Okay. And having said that, I let myself digress. Having said that, they also say "That's not the appropriate way to truly measure complications." So for multiple of these measures, the societies said "These are not the right things to measure." And again, our

purpose is not to define what is the right measure at this time, only to comment on these things. But these were several of the things that were mentioned in that letter.

CO-CHAIR BASKIN: And I think we should open it up for comments here from the group, if there's a little bit of discussion regarding this exception question.

CO-CHAIR SAIGAL: Just that the exception question is supposed to be a true exception, that we think there's a very serious, compelling reason. There shouldn't be any negatives, in my view, that prohibit us.

CO-CHAIR BASKIN: And I think what I'm essentially hearing, or what I think from the prior conversation, is that some of these components could potentially fit into a quality index, quality score, quality composite, but as a whole that it doesn't seem to meet the level that we would want for an exception, as for this whole composite, but

that there's potentially some feedback here for the developer.

Is there any other comment before we go on to have this vote?

(No response.)

CO-CHAIR BASKIN: Okay. So 1 is voting yes, there is an exception, a compelling reason to move this forward. 2 means no, this will not be an exception, the evidence criterion has not been met.

So, those voting yes, 1, please raise their hands?

(No hands.)

CO-CHAIR BASKIN: So that is zero for yes. So 2, no exception being made?

(Show of hands.)

CO-CHAIR BASKIN: And that appears to be unanimous at 15. Okay. So the exception criterion is not there. Do we even discuss the gap at this point? We don't need to go there. And then, of course, the measure, we can't move forward, so we don't

vote for that.

So at this point, I would say any comments, any feedback for the developer, this would be a good chance to do so. I think my comment just a few minutes ago was one for the developer, that there may be some reason to believe there's evidence that some of these components may be very meaningful in a quality index, and that one potentially could create one for which all the components would be acceptable to this group, but that at this time, this one does not meet that level.

But I would appreciate others to comment as well. And Zahid, do you want to make a comment to the developer?

MEMBER BUTT: Yes. I'll just again state that this type of all-or-none type of quality index, which weights things that are on the one extreme serious complications the same as bowel prep, I think should be looked at again. Because I think, really, there is need for an indicator like this, but

probably more like a true composite, where there is some actual scoring, perhaps with weighted scoring of each component, and perhaps the components, maybe two or three of those could be consolidated into a single component.

CO-CHAIR BASKIN: Well, I appreciate that. I think you're looking down the line in terms of implementing this and the impact of the measure itself, not the impact of the concept. But from the point of importance, how it would be scored or methodology is probably not one of the components that we would vote at this level of Phase 1. But I appreciate that feedback.

MEMBER BUTT: And I fully support the consumer's desire to have that single number, that they can say whether this is a good number or not, that truly represents the procedure itself.

CO-CHAIR BASKIN: Any other comments that someone wants to have? Anne?

MEMBER PELLETIER-CAMERON: Just a quick comment about the length of this document. I just think the rest of these measures today that we've measured have been documents of reasonable size, and this is a 115 page document, and most of it is not necessarily even focused on the form and format. I think it's a little bit difficult for members of this panel to go through that much volume. I think it speaks poorly of this presentation that it's 115 pages, 20 of which are a PowerPoint at the end.

CO-CHAIR BASKIN: I understand that some information may have been provided that was beyond the scope of the Phase 1 part of this, which is just the section we were to talk about. And yes, I appreciate that the PowerPoint provided information regarding performance gaps, but it went beyond that to some information that just wasn't necessary for this committee. I appreciate that comment, that we need to filter down this

information to make it a little more -- make all of it pertinent to the decisions that we're making.

MEMBER BORDEIANOU: I just wanted to say that yes, we don't rate the scoring, but I think that there is a standardized scientific way of developing indices that go through a validation process, and the Delphi process, and the societies that sent this lovely three page letter could perhaps unite and create an index that could then be used.

CO-CHAIR BASKIN: Yes. And I'm not saying scoring is not important. I'm just saying that it would have been discussed at the second phase, and not at this phase, in this context. That's all.

MEMBER FALLER: Out of deference to the consumer and the gastroenterologist, I think the car analogy's great, but I want to know whether the brakes are going to fail, and not whether the valve isn't right.

CO-CHAIR BASKIN: Jenifer?

MEMBER LIGHTDALE: Actually, I'd follow up on that. I mean, at the end of the day, this is intended to be a score that a patient would use to say that it's okay to go to this gastroenterologist. And there, actually coming back to this question of process and outcomes being a hybrid, like here, there's only one outcome in here, and it's not adequate to really say that you'd be safe, that the brakes would not have failed, because the endoscopist would score very highly on not having immediate complications, but we don't know what their late complication rate is.

CO-CHAIR BASKIN: Yes, and I do think that was a difficulty, and the fact that there was a little bit of a mix of process and outcomes, and that made it a little bit harder to wrap your hands around this.

MEMBER LIGHTDALE: And getting back to the whole ADR question, throwing in adenoma detection rate would be another

outcome, you might start to have a heavier weight on your outcomes when weighting out your process. Or make it all process. It's up to you.

CO-CHAIR BASKIN: All right. Well, thank you everybody for comments, and thank you for a thorough review of that, and thank you to the developer for presenting this. I hope you've got some good feedback. We are beyond our break time, so we're going to take it now, but I do think we should go ahead and take at least a 10 minute break. It's 10:50. If we could reconvene at 11:00, that would be great. Thank you.

(Whereupon, the above-entitled meeting went off the record at 10:50 a.m., and was resumed at 11:02 a.m.)

CO-CHAIR BASKIN: Let's take our seats and get started. We have some very anxious presenters, I'm sure.

Okay. We have three more measures to consider, and these measures -- I guess two

of them, the measure stewards are the developers or the AGA, so I guess we'll ask the AGA to make three minutes or so of comments regarding those two measures before we go into the individual reviews. Do we have a representative, or someone on the phone? Is there someone on the phone representing the AGA to present the two measures, 2059 and 2062?

(No response.)

CO-CHAIR BASKIN: Is the phone open? Anyone? They were scheduled for 10:45 and it's 11:04, so they should be there. Was it supposed to be --

MS. ROBIN: Can you hear us?

CO-CHAIR BASKIN: Yes. Hi. Could you say who you are? And I presume you're representing the AGA.

MS. ROBIN: Yes. This is Debbie Robin for the AGA.

CO-CHAIR BASKIN: Oh, hi, Debbie. It's Andy Baskin. If you would present those

two measures, that would be great. Thank you.

DR. BRILL: This is Joel Brill. Yngve will be presenting, and Debbie Robin is also here. We're all on the phone, and Debbie and I will mute so we don't hear all the echoes.

CO-CHAIR BASKIN: Okay. And try and keep it down to about three minutes or so. Thank you.

DR. FALCK-YTTER: Okay. Debbie, do you want me to present this real quick?

MS. ROBIN: Yes, please.

DR. FALCK-YTTER: Okay. So thank you very much for letting us present this. These are the two measure concepts being presented today here to the steering committee for consideration, the NQF C2059 and the C2062. They address the management of the bowel patient with inflammatory bowel disease on long-term corticosteroid therapy.

Both measures are basically intended to raise the provider awareness of

the toxic effects of the long-term corticosteroid use, particular at the greater dose of 10 milligrams a day, so 10 milligrams or greater a day, and so that they can take proactive steps to minimize the dose for those suffering from the IBD diseases that we are talking about.

So these two measures basically are preventive care measures, corticosteroid sparing therapy, and the other measure, corticosteroid related iatrogenic injury - bone loss assessment, so it's an assessment measure. And they are part of the 2012 PQRS inflammatory bowel disease measure group, and it's also a proposed PQRS measure for 2013.

These measures were developed during 2010 and 2011 by the AGA, utilizing the PCPI independent measure development process. The multi-stakeholder workgroup included representatives from the Crohn's and Colitis Foundation of America and the American Society of Colorectal Surgeons. The workgroup was

co-chaired, I believe, by John Allen, from Minnesota Gastroenterology Group and the University of Minnesota, and Themis Dassopoulos, and he's currently at the Washington University in St. Louis.

I'm currently the lead methodologist for systematic review being conducted by the AGA on immunomodulators and biologics for moderate to severe Crohn's disease. I am Chief of GI here at Cleveland VA Medical Center. I am at Case Western Reserve University. If you have any questions in regard to that systematic review and the evidence supporting these measure concepts, I will be happy to go into much more detail.

Dr. Brill, as we just heard, is also on the telephone right now, and he has also supported this measure group. And Debbie is also on to tell us about some other details if necessary.

So, just a few words on the background of this. Approximately 40 percent

of patients with inflammatory bowel disease are treated with corticosteroids. The initial therapy with steroids is associated with much poorer prognosis, including inability to taper off the steroids without experiencing flare-ups of disease, and disabling symptoms and surgery. Some of the population-based studies have shown that steroid dependence occurs in one third of all the IBD patients treated.

As you all know, major steroid-related side effects for adult patients with Crohn's Disease are metabolic bone disease and infectious complications. There's a risk of comorbidity, including sepsis and fractures associated with long-term high-dose corticosteroid use. Patients with Crohn's Disease and UC are at an increase risk of death during the periods of current corticosteroid use, while treatment with thiopurines have not been associated with an increased risk of death.

Increasing the treatment of IBD patients to steroid-sparing drugs, the use of dependency on corticosteroids decreases, along with the risk of comorbidities. The increased risk of infections is probably attributed to the disease's severity, also, but concomitant steroid use probably plays a larger role. The use of prednisone is a strong independent risk factor for serious infections and death.

So we use steroid-sparing drugs. These are immune suppressant biologics that can provide us alternatives to treating with corticosteroids alone. I think that's important. And introduction of those agents into the IBD treatment regimen provides the opportunity to minimize those exposures to corticosteroids and their side effects. Despite the advantage in the therapy of IBD, considerable subsets are still kept on prolonged steroid therapy.

Comprehensive literary review and analysis showed that, although the majority of

patients with active Crohn's Disease respond rapidly to steroids, about half will be either steroid-resistant or steroid-dependent in one year.

Osteoporosis in itself is recognized as a complication from IBD and steroid therapy, and it contributes to the increased risk of osteoporosis observed in IBD. Long-term steroid uses are associated with an osteoporotic fracture rate of 30 to 50 percent, mostly at the sites of the vertebrae, hips and pelvis. And to minimize bone loss by using alternate therapies, alternate steroid therapy has actually failed to reduce those fracture rates. So it's really not something that we can do in practice.

In a population-based study from the U.K., they've cited an unadjusted relative risk of hip fractures of 1.62 for IBD and 1.49 for UC, and 2 for Crohn's Disease. So about twice as high, the risk to the general population.

Wagner et al. performed the survey inquiring into the awareness of implementation of the IBD or AGA guidelines on osteoporosis in IBD patients. Slightly less than half of these respondents used these guidelines for decision making in the management of IBD patients. So physicians who are self-reported utilizing these guidelines adhere to those recommendations.

There were other studies conducted by Wagner and others that have shown disparity by rating insurance status in the management of IBD, racial and socioeconomic disparities have been identified in osteoporosis screening and treatment. Details of these studies and their findings are included in the submission material.

The AGA, which is in the process of conducting a systematic review of this issue, thanks the National Quality Forum and steering committee for the opportunity to present these measure concepts and to

participate in the redesign of the endorsement process. Thank you.

CO-CHAIR BASKIN: Thank you for introducing the measures to us, and for preparing that. Our presenter is Zahid. Go for it, impact.

MEMBER BUTT: Thank you. Yes, this measure demonstrates that gastroenterologists do take care of patients, not just scope them.

(Laughter.)

MEMBER BUTT: I do have a couple of questions for the developers, if I may have your permission to ask a couple of questions.

CO-CHAIR BASKIN: Go ahead.

MEMBER BUTT: Okay. So my first question is that, under the denominator exclusions, you have a statement at the end that says "We have been able to include a patient exclusion, for example if the patient refuses steroid therapy," but you also exclude patients who are not on steroid therapy.

What's the difference between those two?

DR. FALCK-YTTER: Debbie, maybe you want to answer that, because that was confusing to me too. But that basically means -- it's a technical issue, right, Debbie?

MS. ROBIN: Yes. This has to do with the way that we have had to struggle with the coding when this measure was initially developed in terms of PQRS, and thinking about administrative coding. We have since had the ability, and have been exploring use of this measure through our recognition program, which is a registry-based program.

The point of that comment was simply to say that we have been able to better incorporate exclusions and have some flexibility around the actual denominators and the exclusions that we did not have when we originally developed it in PQRS.

CO-CHAIR BASKIN: I mean, we can talk during the comment section at the end of these comments whether we think the exclusions

are appropriate or inappropriate, but I think we'll get to that after the review.

You had another question, though, Zahid, for them?

MEMBER BUTT: Yes. And one quick question. I just wanted to make sure that the denominator statement is just people who are age 18 and have inflammatory bowel disease, and the denominator does not include "who have been on steroids." You include that in the numerator?

MS. ROBIN: Yes, we have -- again, the way it's currently coded, that is -- for PQRS and administrative purposes, there's a separate numerator to identify patients who are not on that level of corticosteroid treatment. Again, we had hoped to have some more flexibility there, but we're limited by the current coding requirements at the time the measure was developed and went to PQRS.

MEMBER BUTT: Okay. Thank you. So this is a measure that tries to take

patients who are on corticosteroids as defined by the measure: prednisone 10 milligrams or equivalent -- and there's a little table that they have for that -- who have been on this does for 60 or greater consecutive days and have been prescribed corticosteroid sparing therapy, as in immunomodulators, such as imuran or 6-MP, or the biologics, the anti-TNF agents.

The denominator, as mentioned earlier, are all adult patients with a diagnosis of inflammatory bowel disease, both ulcerative colitis and Crohn's Disease. The data source is electronic clinic data, registry data, and the level of analysis is at the individual physician level. I assume that claims could also qualify as a data source the way the numerator is being captured, and certainly the codes for the denominator would be there, although it's not specifically mentioned in the submission.

So in terms of its impact, there

is a significant body of evidence that's presented that inflammatory bowel disease is a fairly common disease that is treated by gastroenterologists, that 40 percent or so in one study patients with IBD will require longer-term steroids, and it is sometimes difficult to get the doses below the dose that's considered to be a relatively high chronic steroid does.

And certainly, prolonged steroid exposure, there is data to suggest that it is associated with several potential complications and side effects. There is also a body of evidence that the steroid sparing agents, when used, do not have the same level of problems and complications, and are at least as, if not more, effective than the chronic use of steroids.

So, based on the evidence that's presented, it appears that this should be a high-impact condition in my opinion. I'm certainly interested in seeing what the others

think about it.

MEMBER MORTON: I definitely agree, this is high impact. And I think there's been a lot of mention about some of the issues around bone necrosis, but one thing that should be brought up in terms of a complication is iatrogenic obesity. There are so many patients who are on steroids who gain so much weight, and if they end up seeking therapy, like say bariatric surgery, it's very complicating. So I think this measure is very, very important.

MEMBER PELLETIER-CAMERON: This is more of a question than a comment. Not being a gastroenterologist, why would someone prescribe steroids to a patient with this body of evidence? Is it lack of knowledge, or is it cost to the patient?

MEMBER BUTT: I think that, generally speaking, not everybody who has inflammatory bowel disease initially will require long-term steroids. So steroids often

can be used on a short-term basis. You can get the patient off. Many people don't have a recurrence.

So it is an accepted form of initial therapy. The issue is really the chronic use of steroids, and it is that subset of patients that you can't get them below a certain dose, or can't get them off of it over a longer period of time. That's where the complications come in. So it's generally reasonably safe in the short term, but the complications are more problematic on a long-term basis.

CO-CHAIR BASKIN: Liliana, go ahead.

MEMBER BORDEIANOU: Just a comment. Essentially, it's an issue of maintenance versus induction of remission. Steroids are used in an acute setting to induce remission. They're very effective. But there are other medicines to then maintain patients in remission with less complication

profile, and physicians focus on the acute and forget about the follow-up. And that's what this measure is getting at.

MEMBER LIGHTDALE: That's exactly right, so that's key, but the other thing is this metric is all about chronic use of corticosteroids, but it doesn't tackle -- and I did put it in my comments, and I don't know if this is the right dramatic moment to bring it out --

(Laughter.)

MEMBER LIGHTDALE: -- but it really doesn't actually balance out with -- okay, so you get them off corticosteroids onto these immunomodulators, biologics. They come with a whole host of side effects, a whole host of other issues.

And that's the only trick here, is going to be, how do you do this in a way that you emphasize the importance of avoiding inappropriate chronic use of steroids? Because not everybody responds to the other

agents, so that's another reason you can wind up on chronic steroids. But anyway, how do you avoid inappropriate use and not actually push it to the point that somebody says I could treat you in my urology practice for your IBD, and start you right away on 6-MP and mess things up?

MEMBER BUTT: I thought your dramatic statement was going to be "Why is it just for adults and not pediatric patients as well?" Because in them, it's an even more important issue, in some cases.

MEMBER LIGHTDALE: Actually, I think, honestly, we should be exploring a little bit whether this should be restricted to over 18. The same issue is going on right now in pediatrics, and I'm not sure if that's coming from the Crohn's and Colitis Foundation, to keep it to greater than 18, if it's just my adult GI colleagues who are being respectful of pediatrics. But I would encourage us to think about that.

CO-CHAIR BASKIN: I think we can just accept that now as a comment to the developers, so we don't forget it.

MEMBER BUTT: I think her comment actually is addressed in the body of evidence, later on. Because they do present some evidence that the anti-TNF in aggregate have less side effects than steroids, for whatever that's worth. And I think there is a long body of evidence for the immunomodulators, that in that context, where you have to keep people on very high doses of steroids on a long-term basis, that the immunomodulator has less aggregate harm than chronic high-dose steroids. I think that would be --

CO-CHAIR BASKIN: Let's get the impact part out of the way, because I think we're getting into the evidence and the quality of the evidence.

So, specifically around impact, what I've heard so far is that IBD, fairly common disease. It's certainly a serious

disease. The treatment, chronic steroids, certainly serious. It has significant side effects. So, impactful in terms of the severity of the complications involved, and that there is alternative, less complicating treatment. There is alternative therapy which is preferable, to at least reduce the dose of steroids or eliminate the steroids when possible, and there's a sizable number of IBD patients who are in this situation with chronic steroids, to cause the impact. So -- if it's directly to impact, Robert, then please, go ahead.

MEMBER ELLIS: Just quickly, can any of you quantify for me what a fairly common disease means in the U.S.?

MEMBER BUTT: I don't know about the numbers, but it is probably the second most common condition that GI treat, after GERD, right? What would you say the total numbers would be?

MEMBER SCHOENFELD: It's at least

a -- the U.S. population is 275 million.

CO-CHAIR BASKIN: It's three hundred and something, but you're close.

MEMBER SCHOENFELD: So I know the estimate is 1 out of 300 people have inflammatory bowel disease, so we're pretty close to a million on this, then.

CO-CHAIR BASKIN: So it's not just volume, but it's not a rare disease by any means. But the impact on those that have the disease is fairly significant, and that's one way to measure impact as well.

MEMBER LIGHTDALE: This is from the CDC website: "1.4 million persons in the United States."

CO-CHAIR BASKIN: Any other comments regarding impact, or we'll go to a vote on impact? Zahid, mic off when you're not speaking. And Jenifer as well.

So let's go to a vote, then. High, moderate, low, or insufficient. So high, voting 1 for high?

(Show of hands.)

CO-CHAIR BASKIN: Okay. So that's 14, and we only have 14, because Chris had to leave early. So then, obviously, zero for moderate, zero for low, and zero for insufficient evidence for impact.

So now, let's move on to the quality of the evidence involved.

MEMBER BUTT: So in the evidence, they do present several studies, and also refer to guidelines. In one of the AGA institute guidelines, there is a grading. Grade A is assigned to where long-term treatment of corticosteroids is undesirable and patients with chronic, active corticosteroid-dependent disease should be treated with immunomodulators. There's another reference to a Crohn's Disease and UC study for immunomodulators that has been graded as a C.

In terms of the risk profile that was mentioned earlier, there is moderate to

high certainty, in the estimate of the quality of evidence, that the use of immunomodulators and/or anti-TNF is effective in inducing and maintaining remission in IBD to the degree that patients can successfully taper off the steroids.

And then there is also the overall body of evidence regarding the use of immunomodulators for steroid-free or steroid-taper remission, which includes five randomized controlled trials that looked at failure to achieve remission, and two randomized controlled trials that were aimed at examining disease relapse. The overall body of these RCTs was moderate, and there was no significant risk.

Same thing with the anti-TNF. The overall body of evidence for the use of anti-TNF agents in inducing and maintaining remission, allowing for successful taper and steroid-free treatment is moderate to high. And as was mentioned in terms of the harm,

there is evidence to suggest that there is a balance between benefits versus harm which favors the use of these agents compared to long-term steroid use.

So I would say that, on balance, the body of evidence would be somewhere between moderate to high.

CO-CHAIR BASKIN: Comments from others? It does seem like there's actually, for one of the few times, direct evidence of actual comparisons and some randomized controlled studies that specifically speak to the issue, which is unusual for us to have. But our gastroenterologists may think otherwise.

MEMBER LIGHTDALE: I know more about IBD than I care to admit. So the only issue is -- of course, this is thinking about an entire global body of evidence, and we're not really talking about effect size, or also exactly what the outcome was of each of these trials. And many of them were kind of

short-term, and didn't look at long-term, that kind of issue. So they're heterogeneous in that way.

MR. AMIN: Andy, the only question I have for you and the group is, the moderate to high is across quantity, quality, and consistency of the evidence that was presented?

MEMBER BUTT: Yes. Consistency is high, quality would be moderate, and quantity I would say is high.

CO-CHAIR BASKIN: Any comments to that classification there?

(No response.)

CO-CHAIR BASKIN: There seems to be some comfort with that. Any more discussion before we vote regarding the evidence?

MEMBER BUTT: I think, just to address Jenifer's comment, the studies are short because these are relatively new therapies. So by definition, we don't have

long-term data. In the immunomodulator population, we do have long-term data, but the biologics are relatively new.

CO-CHAIR BASKIN: Yes. I actually was interested in that some of the data, while it would seem fairly recent, because it goes up to 2005-2006, in the world of treatment for IBD, when these treatments are -- the curve is pretty steep in terms of the utilization of these things. Is this even valid today, five or six years later? I'm not so sure the performance gap exists the same way as it does. But we haven't gotten there yet. I just don't know that the problem is the same problem anymore.

MEMBER LIGHTDALE: To comment to that, the one drug we've had around for a very long time is steroids, which is why it's cheap and there's so much data on it. And frankly, everybody's starting to look now at the adverse outcomes, as opposed to "Thankfully, there was a drug to help with IBD." And so

the bottom line is you don't want people practicing old-school IBD care and leaving somebody on steroids for a long time.

CO-CHAIR BASKIN: So, let's bring this to a vote. Our options are 1, 2, or 3. 1, yes, the body of evidence meets our criteria, 2, it does not and it does not exist, 3, insufficient submitted, but the body of evidence is out there somewhere.

So, those voting 1, please raise their hands.

(Show of hands.)

CO-CHAIR BASKIN: And that appears to be unanimous. Is that a 14 count? Okay, which obviously means no one's voting 2 and no one's voting for 3.

Okay. Let's move on, then, to the performance gap.

MEMBER BUTT: So in terms of the performance gap, in the opportunity for improvement section there is some information that's provided. There is a study that shows

that -- it's a relatively small study, that does show that there is a performance gap. It appears that there isn't a lot of studies that were presented in this proposal that show or say a performance gap exists based on the presented information.

But I think, in my sort of own small sample of 15 gastroenterologists in one practice, I can tell you from experience that there is a significant variation in the care that's delivered. And I don't know if that qualifies for this type of evidence, but I don't know if there is any additional studies or data that could be presented that show a performance gap exists. But my guess is that it does exist.

CO-CHAIR BASKIN: John, you wanted to comment?

MEMBER MORTON: I would say, having looked at some of these measures through yesterday and today, I think these guys did a very good job in documenting the

performance gap. I think they were the only ones who actually addressed disparities, for that matter. So I think they did an excellent job in assessing this, and indicating there was a performance gap.

CO-CHAIR BASKIN: It doesn't seem like there's a volume of evidence regarding performance. Am I missing something?

MEMBER MORTON: I'm going by the disparities, more than anything else.

MEMBER BUTT: There's more data on the disparities section, yes.

CO-CHAIR BASKIN: Jenifer, did you have a comment, or your thing is just up? Your card's still up, just because it's still up. Okay.

Any other comments regarding performance gap?

(No response.)

CO-CHAIR BASKIN: So I guess what I'm hearing is that --

MEMBER BUTT: It's probably

moderate, I would say.

CO-CHAIR BASKIN: Yes. One could say that if there's data in the disparities section that there's a significant performance gap -- I mean, that in and of itself may show moderate to high performance gap, whether it's in the general population or not.

MEMBER MORTON: Well, I think the performance gap is particularly specific to racial disparities, because some of these drugs aren't routinely available to patients with lower socioeconomic status, because they are of higher cost, and there's a lot of drug decision panels about who should get them. So I think the gap is really prevalent there.

CO-CHAIR BASKIN: Unless there's other comment, then I think we can come to a vote on performance gap. Once again, four choices: high, moderate, low and insufficient evidence.

So, let's vote. 1, high. How many vote for high?

(No hands.)

CO-CHAIR BASKIN: And actually no votes for high. So how many voting for 2, moderate?

(Show of hands.)

CO-CHAIR BASKIN: And I think we have 13. Thirteen for moderate. How many voting for low?

(Show of hands.)

CO-CHAIR BASKIN: We have one vote for low. And I presume, then, there's no votes for insufficient, since that adds up to 14. Okay, so we made it through the performance gap. Now, I guess, we move on to recommending this concept.

Any particular comments somebody wants to make regarding this? Okay, Liliana, you go first, then.

MEMBER BORDEIANOU: I'm sorry, it's the surgeon speaking, but one of the corticosteroid sparing therapies is a consult with a surgeon, because that's another

treatment in getting people off steroids. So maybe that could be included under the exclusions.

I'm saying that the list, they're measuring who was prescribed anti-TNFs, methotrexate, et cetera. But the other thing that might have happened is that the patient was referred to a surgeon for discussion about surgery, and I don't know if they're capturing that. I guess that goes more to how it's being measured than whether or not it should be approved or not, so we could discuss more in the second phase.

CO-CHAIR BASKIN: Go ahead, John.

MEMBER MORTON: I think that's an excellent point. That's actually one of the indications to do a total abdominal colectomy. So it's probably more meant for feedback for the developer, but I think it's a terrific point.

CO-CHAIR BASKIN: I still have a struggle with this numerator and denominator.

I guess maybe I'm misunderstanding here, because it seems to me that the denominator is all patients with IBD, treatment or no treatment. And the numerator, to get a hit in the numerator, you have to be on long-term steroids and be on a sparing agent -- which is a good thing -- but what about the people who are on an anti-TNF factor who aren't on steroids at all? You don't get credit for that as being a good thing? I guess I don't understand how this differentiates good and bad care.

MEMBER BUTT: I was saving some of that commentary for last, but I wasn't sure where to plug that in. And that was my original question. Really, for this measure to be really effective, the denominator should have been patients who are with IBD and have been on chronic steroids. And of that percentage, what percentage were then prescribed anti-TNF therapy? Because I think the hole in this measure is that it misses

those that are on steroids and were not prescribed anti-TNF therapy.

So in other words, if you look at your pie of your denominator as all IBD patients, and you take another circle the patients who are on steroids, this takes a slice of that, those that were prescribed anti-TNF therapy, but then it takes that as a numerator and assigns it to the IBD as a denominator, and it kind of loses some of its fidelity there.

CO-CHAIR BASKIN: Jenifer?

MEMBER LIGHTDALE: I actually struggled with this with the next one, as I was trying to understand it. I think the reason that they wrote it this way -- and it's not well-written -- is because of these CPT II codes that they're using.

And so basically what they said happens, because they're only assigning this -- and I think this is what the person before was trying to explain. But because they're

only assigning that -- I think, and you guys can tell me if I'm wrong -- they're only assigning CPT II codes to somebody who's 18 years or older with a diagnosis of IBD and who's on steroids, and that's essentially your denominator. Like, they've sort of artificially written it in a way that reflects their coding.

CO-CHAIR BASKIN: If that's the denominator, then the description should be that that's the denominator. So maybe we can ask the developers.

MEMBER LIGHTDALE: Can I follow myself up with one quick thing?

CO-CHAIR BASKIN: Please.

MEMBER LIGHTDALE: A simpler question is whether it's simply patients with IBD who are managed with corticosteroids for greater than 60 days over all patients with IBD. I mean, that's all -- never mind the steroid sparing agent. That's how you get them off the steroids. You just don't want

them on steroids forever, so why not just make it about being on steroids for greater than 60 days over IBD?

MEMBER BUTT: Can I make a comment? So I think this was the limitation of -- you're probably correct -- the CPT II. Because what you'd have to do then, is you'd have to assign a CPT II code to all of your denominator cases that have IBD and are on steroids, and are on steroids greater than 60 days. So you would have had to assign a lot more CPT IIs. So they tried to sort of reduce the burden, but within that process, I think it lost some of its value.

CO-CHAIR BASKIN: Well, I'm going to ask the developer here to jump in again, because I'm still confused as to who's in the denominator and who's counted as a numerator hit, meaning a positive hit, like you did the right thing and you get credit for it.

So explain once again the population of the denominator. Is it -- it

says here just "those over 18 with a diagnosis of inflammatory bowel disease." Is that the true denominator, anybody with inflammatory bowel disease over 18?

MEMBER BUTT: I think they said yes when I asked that question.

CO-CHAIR BASKIN: Folks, are you out there?

MS. ROBIN: This is Debbie Robin again. I will address this as succinctly as I can. The denominator for purposes of PQRS is defined by diagnosis and service codes. There is no combination of those elements currently available that identifies patients with IBD who are on chronic corticosteroid treatment. Therefore, what we had to work with was to use existing codes that allowed us to identify all IBD patients.

Then, for the various measures or calculations, we then developed various CPT II codes. There is a specific code that identifies patients who are not on long-term

corticosteroid therapy. So from a performance perspective, there's a way to calculate it so that those patients are taken out of the equation.

To allow people to report this measure, we had to sort of be able to find a way to allow them to report it in that manner with the limitations of the diagnosis codes. Having said that, in an ideal world with electronic specifications, which we do plan to get to in the future, is that yes, we would create the ability to pull out just those patients who are currently being treated with long-term corticosteroid therapies, and that would be the denominator.

CO-CHAIR BASKIN: But essentially that's what you've done, then. You've just done it by saying "Take all the inflammatory bowel disease members. Those with the CPT code that says they're not on chronic steroids are excluded, so theoretically what remains is those that are on chronic steroid therapy. Is

that essentially what you've done?

MEMBER BUTT: No. I think the way she explained it is that you have to use two separate CPT II codes in the numerator. One would capture the ones that are on chronic steroids and receive the steroid sparing. The other would be the ones who are only on chronic, and you would have to take the two rates together to come up with the answer to the single question that we were asking originally.

MS. ROBIN: Yes, that is correct.

MEMBER BUTT: But they also say that, in the electronic specification, the denominator definition will change.

CO-CHAIR BASKIN: Were you going to say something?

MEMBER BUTT: Is that correct?

MS. ROBIN: That's correct.

DR. BRILL: Yes, that's correct, Andy.

CO-CHAIR BASKIN: When we go to

feasibility in the next level we'll see, but if it says what it's measuring I'm okay with it. I'm just not so sure it is. But I think you have found a way to do it, so I'll let it rest.

MEMBER BUTT: Can I make one final comment, then? Potentially, as long as this remains a measure and it is not replaced by the electronic measure, perhaps it should be a paired measure, where you include the other one as well?

CO-CHAIR BASKIN: Well, let's first vote on approving this concept to move along or not, and then any comments to the developers can make.

So, any other discussion before we come to a vote on this?

(No response.)

CO-CHAIR BASKIN: Okay. Let's come to a vote, then. Those saying yes, recommend approval of the concept?

(Show of hands.)

CO-CHAIR BASKIN: That appears to be unanimous. I'm guessing that's 14 people. So that means zero noes, okay.

So, any additional comments for the developer?

MEMBER BUTT: So, that was the comment that it should, perhaps, be considered as a paired measure with the other one, with the second CPT II, as long as this will remain in circulation.

MS. WILBON: Are you talking about the measure that we're getting ready to discuss next?

MEMBER BUTT: No. What I'm saying is that there is a CPT II code related to this that captures the patients who are on chronic steroids, but have not received anti-TNF therapy. That percentage calculation, in combination with this, would give us the answer of what percentage of patients on chronic steroids were put on anti-TNF or immunomodulative therapy.

I hope I'm not confusing people.

CO-CHAIR BASKIN: So I think the comment is basically to the extent that the CPT codes can help in measuring this in a simpler way, that that would be advantageous for all.

Okay. Any other comments for the developers before we move on to the next measure?

MR. AMIN: I guess the only question I have, Andy, is it sounds like there are some questions here related to the construction of the measure.

CO-CHAIR BASKIN: Right.

MR. AMIN: And to the extent that we can be as specific as possible on what you would expect to see when this measure comes back in Stage 2, if there are some changes related to the construction of the measure, the more specific we can be there, the better.

CO-CHAIR BASKIN: Well, my comment is simply that when you look at the

denominator statement, it says "all patients with inflammatory bowel disease." And if you have, in fact, excluded a large group of patients with inflammatory bowel disease, and it's a significant exclusion, then the denominator statement's really not accurate. It's really not all patients with inflammatory bowel disease.

If there are so many exclusions, that should be part of the denominator statement, so it's very clear who is left in the denominator. That's my only point, is that when there are exclusions that exclude 2 percent of the patients, they can be exclusions. But if it's an exclusion that excludes a large percentage of the inflammatory bowel disease patients, then it's part of the denominator statement, to me, in terms of a reader who's trying to understand what a measure says.

MEMBER BUTT: But what exclusions are you referring to?

CO-CHAIR BASKIN: There's an exclusion here that says "Because of the use of clinical data, those that have not received a dose of corticosteroids greater than or equal to [...] are excluded from the denominator." It seems to me that that's a large patient population. If they're excluded, they're excluded.

MEMBER BUTT: But in the ideal world, the patients that they are looking for in the denominator are those that have IBD and are on chronic steroid therapy. So by definition, those who are not on chronic steroid therapy would be excluded.

CO-CHAIR BASKIN: Right. And if that's a large group of people, then that should be reflected in the description of the denominator. Don't call the denominator "all patients with IBD" when 25 percent of the people with IBD, or 50 percent of them, aren't included in the denominator. Call it what it is.

MEMBER BUTT: I see. That's really what that whole discussion was about --

CO-CHAIR BASKIN: Yes, I understand

MEMBER BUTT: -- that in the ideal case, the denominator statement should be different. It should include "and those who have been on chronic steroids."

CO-CHAIR BASKIN: Yes. So we're just saying it in two different ways.

MEMBER BUTT: Right. It is the constraints that they have, so they have to go with the CPT II code within the numerator.

CO-CHAIR BASKIN: That's an implementation issue. The description of the denominator is not how you got there, it's what is -- a reasonable person looking at this is going to look at a numerator and a denominator, and they should reasonably be able to tell what we're measuring. And I'm having trouble telling that from this description.

MEMBER BUTT: Right.

CO-CHAIR BASKIN: How they got there is all in the behind-the-scenes stuff.

MEMBER BUTT: Right.

CO-CHAIR BASKIN: John, did you want to make a comment?

MEMBER MORTON: I was just going to say if the question is how many patients out there are diagnosed who don't get therapy, I think that's probably a pretty low number. And I mean, my GI colleagues can comment on that, but I would think that would be a pretty low number.

CO-CHAIR BASKIN: Yes, but I was speaking to the exclusion group is the folks that are theoretically not receiving chronic steroid therapy. That actually may be a sizable number. I'm not talking about people who aren't treated at all. The exclusion is for people who are not being treated with chronic steroids. That could be a decent group of people.

MEMBER BUTT: But I think the key group that you want to get at is the ones who are on chronic steroid therapy but did not get anti-TNF. And this measure construct does not allow you to do that. The only way to do that in the current CPT II framework is to assign the second CPT code, which says that this patient did not meet this criterion of having transitioned to it, but out of the IBD patients, they were on chronic steroid therapy.

So that's the second CPT, and that's where my recommendation was, that as long as this is going to stay, that maybe they should include the other one as a paired measure, so that the two of them combined, one will tell you the percentage of people who were on chronic steroid therapy but did not get anti-TNF. This one would tell you the percentage of IBD patients who were on chronic steroid therapy and received the anti-TNF.

CO-CHAIR BASKIN: Okay. We need

to pull this one to a close, because we've really fallen behind. So if you've got 15 second comments, you can make them.

MEMBER LIGHTDALE: The 15 second comment is, I guess the only thing I'd be advising is, first off, generalize it, simplify it. And really, the goal here is steroid sparing. I think that has to be key. And that gets to your surgery discussion, too. This isn't just driving people to another drug, this is get them off steroids.

CO-CHAIR BASKIN: Right. And Liliana, 15 seconds or less.

MEMBER BORDEIANOU: Right. The feasibility discussion in the next phase should include how they propose to measure patients that refuse treatment. How are they going to do that?

CO-CHAIR BASKIN: Okay. So thank you, and I guess I'm surprised that took as long as it took. I wasn't watching the time as well as I should have, so I apologize.

So let's move on to the next measure, which is the bone loss assessment. And Jenifer, you're going to present this one?

MEMBER LIGHTDALE: Yes. I'll try to avoid redundancy.

So this was, again, a process measure, and it does again involve these CPT II codes. So the numerator is patients with IBD who have received corticosteroids at least at a threshold does of ten mgs per day for 60 consecutive days who have been assessed for bone loss -- again, all up there in the numerator. And the denominator is all patients with IBD. The level of analysis is a the clinician level.

And in terms of the high impact, the bottom line is both IBD and, independently, corticosteroid use are associated with osteopenia. And if you put the two things together, there's clearly an association with the relative risk of hip fracture going up in patients who have IBD and

are on corticosteroids.

And basically for their evidence, they had two population-based studies. One is from the U.K., from 2004 --

CO-CHAIR BASKIN: Let's just get to the impact.

MEMBER LIGHTDALE: Oh, this is impact. Sorry.

So for their impact, the evidence that they were citing was two population-based studies. And again, for me, they were just older studies, and neither one was in the U.S. So U.K., 2004, and one in Canada in 2003.

CO-CHAIR BASKIN: In terms of the quality of the studies regarding impact, U.K. and Canada, that's perhaps acceptable to us. But did it show a reasonable impact?

MEMBER LIGHTDALE: Again, both studies show that there's an independent risk of IBD for hip fracture and for corticosteroid use and hip fracture. And corticosteroid use plus IBD does increase your relative risk a

bit.

CO-CHAIR BASKIN: Comments regarding impact?

(No response.)

CO-CHAIR BASKIN: Then let's go -- would you have characterized this measure as high, moderate, or low, in your opinion?

MEMBER LIGHTDALE: In my opinion, it was moderate.

CO-CHAIR BASKIN: Then we'll each vote what's in our hearts.

So we're voting now: 1, 2, 3, or 4. 1 is high impact. Raise your hands.

(No hands.)

CO-CHAIR BASKIN: Zero. 2 is moderate impact. Raise your hands.

(Show of hands.)

CO-CHAIR BASKIN: And that appears to be everyone if I counted correctly. Any low impacts or insufficients?

(No hands.)

CO-CHAIR BASKIN: I didn't think

so. I thought we had 14 there. Okay, so 14 moderate and no high, low, or insufficient. So now we're going to the evidence quantity, quality, and consistency.

MEMBER LIGHTDALE: So for this, basically there were two evidence-based guidelines that were cited. One was developed by the AGA in 2006, and then there's also a guideline that was developed by the American College of Rheumatology in 2010, and both spelled out recommendations for prevention, identification, and treatment of corticosteroid-related osteoporosis. Obviously, the AGA one was specifically looking at inflammatory bowel disease.

And basically, the AGA guideline graded their evidence as an A, suggesting it was consistent, well-designed. Again, probably population-based cohort studies with sufficient power. What was a little intriguing was the ACR guideline used the American College of Cardiology grading system,

and they gave themselves a C, which is indicative of consensus, or expert opinion.

So the newer guideline, which is the ACR guideline, is a consensus opinion statement, although it agrees with the AGA one.

CO-CHAIR BASKIN: Were the AGA guidelines specific to IBD patients? Because I think the ACR guideline was not necessarily specific to IBD patients, but just those who were on chronic steroid therapy for whatever reason, presumably a rheumatologic reason, but nevertheless for whatever reason.

MEMBER LIGHTDALE: Full disclosure is, I read what was here. I did not read the AGA guideline. But I do know there are other GI conditions you can treat with long-term steroids, like chronic pancreatitis. There are some others. Autoimmune pancreatitis. So anyway, all by way of saying I think it was mostly focused on IBD.

MEMBER BORDEIANOU: My only

question, and I don't know the answer to that, is we're looking at whether bone loss assessment, i.e. getting a DEXA scan, changes outcomes. And I don't see anything here that suggests that doing the test does anything other than provides you the information that you have osteoporosis, which you could infer if somebody was on steroids for three months.

CO-CHAIR BASKIN: So this is one of those "Is what they're looking for proximate or distal to what we're really looking for," which is treatment, or appropriate treatment based on information.

Are there comments?

(No response.)

CO-CHAIR BASKIN: I don't know whether that's because this was easier to measure, and the other would be much more difficult to measure -- because, frankly, it probably would be much more difficult to measure, because it's not just treatment, there's treatment options, and some of those

treatment options are potentially non-prescription, and some of those treatment options the patient may have potentially chosen not to take, for various reasons.

I guess it gets a little complicated. That's not to say it shouldn't be done.

MEMBER LIGHTDALE: I guess my opinion is probably the ACR was a little bit more careful about being honest that a lot of what they were saying is common sense, and it's consensus as opposed to evidence-based, and that's why you don't have the studies.

MR. AMIN: Just a few follow-up questions. Particularly on the quality, quantity, and consistency, just what your opinion is in terms of what's in here. And also, just keep in mind that consensus-based guidelines would not meet the requirement here, so we could have a discussion around the exception.

CO-CHAIR BASKIN: You or anyone,

so let's just say quantity. Is there a quantity of evidence here, evidence that we would accept, evidence-based studies, to support this testing be performed?

(No response.)

CO-CHAIR BASKIN: And I think he said, at least, that the AGA is, I believe, an evidence-based guideline, that there's certainly a reason to treat these folks, and one could, of course, infer that you can't treat if you didn't test them first. But that's a different level here.

Is there some sense that that's a lot of evidence, or does anybody really know who's here?

MR. AMIN: Well, before we get there, I guess one of the questions here procedurally is that the information that they presented here in the form is very clear in terms of where you would want to vote. So if you want to have a discussion after this vote around what evidence exists, that would be

fine, but it seems pretty clear what information is presented in the form.

CO-CHAIR BASKIN: Well, I'm not sure it's so clear. Because the AGA statement is theoretically an evidence-based guideline, as opposed to the ACR which they're admitting is a consensus-based guideline. So it's not so clear to me that there's not evidence here that is acceptable to us. Whether it's low, moderate or high is, I think, my question.

MEMBER LIGHTDALE: Actually, it was very helpful to have this slide up. So I think there are two good population-based studies upon which the AGA guideline really comes out of, and that would then really qualify it as moderate for quantity. Low moderate, but moderate.

CO-CHAIR BASKIN: So there is at least one thought that there is a moderate amount in terms of quantity. So when we talk about the quality of the evidence, that moderate amount of evidence, those two

studies, randomly controlled studies? Non-randomly controlled studies? Where would you fit those into this construct?

I mean, if anyone else knows, please, feel free. I'm not trying to pick on Jenifer in any way.

Stuart?

MEMBER REYNOLDS: Well, I think it seems fairly insufficient. I mean, it's alluded to in the evidence at the end of the document, but it's not explicitly stated. And so I think I would be comfortable saying it's probably insufficient to evaluate. I mean, it may be that there's data out there that we're not presented with. We may choose to move forward with it even without that data. But if we can't answer these questions, I would say insufficient.

CO-CHAIR BASKIN: Fair statement. Thank you. And John?

MEMBER MORTON: I agree with Jenifer and Stu. This seems a lot lighter in

terms of evidence than what we've seen in the other measure.

CO-CHAIR BASKIN: Okay. Any other comments? And obviously, if there's potentially an insufficient or small amount of evidence, consistency doesn't really come into play. It's hard to be consistent when you're only talking about two potential trials.

Okay. So based upon what I'm hearing here, I think we can come to a vote regarding the evidence that's submitted here. Yes, body of evidence meets our criteria. 2, the evidence doesn't meet the quantity and quality and we don't think it necessarily exists. I think that's going to be a little tricky as to whether that's going to be the case or number 3, insufficient but we think the evidence is out there.

So, think about that for two seconds, and then we'll come to a vote, unless there's any comment that wants to help people who are on the fence, if they're on the fence

between 2 and 3, if anyone wants to make a comment in support one way or the other.

(No response.)

CO-CHAIR BASKIN: I'm not so sure I heard that there is a body of evidence that I can point to out there that exists.

MEMBER LIGHTDALE: What I will be fair about is, I don't think any of us are IBD-ologists. Is that correct? So it's possible.

CO-CHAIR BASKIN: All right. Let's come to our vote, then. A 1 is yes, the body of evidence meets our guidance. Raise your hands.

(Show of hands.)

CO-CHAIR BASKIN: We have one vote for yes. That's all right. We won't state who made that vote.

2, the evidence does not meet the guidance, and we're not necessarily aware that any evidence exists.

(Show of hands.)

CO-CHAIR BASKIN: I think that's five votes. And then 3, insufficient evidence submitted, but we think that body of evidence does exist.

(Show of hands.)

CO-CHAIR BASKIN: Eight. That comes out to 14, right? So 1 yes, 5 noes, and 8 insufficient but the evidence does exist.

MR. AMIN: Insufficient information presented in the form, so the question here is, is there general agreement that the information does exist but it just wasn't presented in the form, and would that body then meet the quality, quantity, and consistency? And if there isn't general agreement, that you can't make this decision at this point, because there's insufficient information, then you would just --

CO-CHAIR BASKIN: Now, wait a minute. I thought the vote went with number 3, that that evidence does exist. I thought this vote only goes if you vote no, which was

the second option, which means that it wasn't presented and we didn't know it existed.

MR. AMIN: So let me go back. Can you go back one second to this? So, maybe I should have clarified this before we voted. If there's a need for a revote, I'm happy to do it.

So, 1 is that it meets. Second is that the evidence does not meet, or that there's no empirical evidence that exists. Third is that there's insufficient information in the form to rate the quality, quantity and consistency, but there is information that exists out there.

So what you do with number 3 is that the information -- since it's insufficient in the form, we ask the committee whether or not there's information that they believe, that there's a body of evidence. And number 2 would be -- just for the sake of completion -- number 2 would be that there's not information that exists, there's not

evidence, but we're going to make an exception here because the benefits outweigh the harms.

So the question here for the group is, is there evidence that you know of that would meet the quality, quantity and consistency? And if not --

CO-CHAIR BASKIN: Okay. So we had eight votes that people thought that that information existed. Now the question is, that information existing, if it had been submitted, would it have met our criteria? But even those that voted 1 or 2 can still vote on this one. It's not just the 8 votes.

But a comment first, before we vote, because this is a little trickier vote.

MEMBER REYNOLDS: Well, I guess the issue that was driving my vote is that, for example, we have two guidelines listed here, and at least one of them is based on evidence, but that evidence is not clearly presented. So it would then lead me to think that that evidence is out there.

I admittedly am not familiar with that evidence. I'm not sure I can, without further discussion, vote one way or the other. But it certainly seems like there's a hint that there's data out there, but we haven't been presented with it.

MR. AMIN: Again, there's a pretty high bar here, just like the exception rule. So the sense would be that the committee would need to put forward that evidence that does exist. And if it doesn't, or it's insufficient at this point, then you would vote no here, that there's not general agreement that it would meet -- it's insufficient, I guess, in this sense.

CO-CHAIR BASKIN: So if you have comfort that you're aware of that information, or comfort that you've accepted others are aware of it, that's fine. You can vote yes here. And if you don't have that comfort level that it exists, or those here that say it exists you're not comfortable that it's

sufficient enough for you, then you would vote no here.

And the result of this, though, would drive us to do what? If we were to vote yes here, then --

MR. AMIN: Then you would move on to gap.

CO-CHAIR BASKIN: Okay.

MR. AMIN: If you vote no here, then the concept stops.

CO-CHAIR BASKIN: The concept stops. Okay. So, let's take it to a vote. Those voting yes, raise their hands. A vote of yes would mean that this could go on to further evaluation.

(No hands.)

CO-CHAIR BASKIN: There are no yeses. And those voting no?

(Show of hands.)

CO-CHAIR BASKIN: Which is unanimous, it appears to be, so that must be 14 of us.

So we're voting no, that there's insufficient evidence provided, and that this committee is not comfortable that a body of evidence exists that would meet our criteria. So then, we stop here.

I think, however, in this particular case, there may be some comments for the developers here, especially if -- and I would say, just off the bat, that if the evidence does exist, this committee would have welcomed it, and just that the expertise in this room is not aware of that body of evidence.

But other comments, please.

DR. FALCK-YTTER: Would you like a comment on that? This is Yngve Falck-Ytter.

CO-CHAIR BASKIN: You know, actually, that's okay. I think we would like to hear it, as long as the comments are short.

DR. FALCK-YTTER: I'll make it very quick. Of course, for full disclosure, I'm a co-developer for the grade system. And

when we make recommendations and these kinds of things, there's a few things to consider. One is, we are not talking about that people who treat IBD patients should sent off patients to DEXA scanning all the time.

It's more about the awareness to actually think about those problems, to have a problem list, and to say "we have thought about and we have assessed that patient," and that it goes into their chart. So it's a very low-effort kind of thing, where people just have to do it.

Now, in terms of how you support this with evidence, it's very clear that this is almost like a good practice point, where you have a beneficial effect in the absence of harm. There's no harm in assessing bone loss. The harm starts when you think that you might actually order a DEXA scan or something like that. So this is only the assessment portion. Every time we have no-harm recommendations, even if the evidence quality is low, you can

still make it a point, make it a performance measure, in my opinion.

But again, these are situations where you have a little bit different way of looking at the quality of the evidence, where you have clearly no direct -- there's no randomized trials that looked at this assessment and see whether they have patient-reported outcomes that are improved. It's just my two cents. I'm sorry to keep you.

CO-CHAIR BASKIN: Thank you for that. Further comments from the group here? Go ahead, Zahid.

MEMBER BUTT: Would that fall in the -- and I hate to use that word -- exception category, then, based on what we've heard?

CO-CHAIR BASKIN: As Taroon was saying, unless we feel that added information has been given to us to make us want to consider exception, we're able to do so. So

I mean, we could certainly talk about it. I don't know that any new information was given to me, other than to say that it sounds like good practice to do an assessment.

And yes, it sounds like it to me, too, but I'm not so sure that's a quality measure, and I'm not so sure that I know what the outcome of that is going to be, how that improves my patient's care. It's not as clearly obvious to me.

But others, please. John?

MEMBER MORTON: I mean, you're invoking a maxim we all employ in medicine, which is you don't order a test unless you can do something with it. Potentially, there can be something done with this. We just haven't seen the evidence for it yet.

MEMBER LIGHTDALE: We haven't discussed whether you could combine this measure into the other one. Are we going to be doing that?

CO-CHAIR BASKIN: Well, that would

be a comment. Let's take the measure in and of itself at this point in time.

And let's be clear, I think it was made clear to us. It's not just ordering a DEXA scan. That does meet the measure. But it's just if you did an assessment and didn't order a DEXA scan, you still get credit on this measure, which is even a different bar, I guess.

Any other comments to be made?

(No response.)

CO-CHAIR BASKIN: I don't see anything compelling here to make us be voting on an exception process here. I don't think that anything has been presented new that would make us do that. So unless I'm hearing a strong voice otherwise, then I don't think that's an appropriate vote.

I'm hearing that. Okay.

Any comments back to the developer, as this measure isn't going forward at this point in time?

(No response.)

CO-CHAIR BASKIN: Okay. I think, though, it's probably -- maybe it goes without saying that if there had been some evidence to show that you can improve the health of these patients, that there would be a better outcome for these patients based on this measure, that would be there. But simply whether you did an assessment or not just doesn't seem to meet that bar.

Zahid, you wanted to make one last comment? I'm sorry.

MEMBER BUTT: Yes. I was just going to say that the previous one will probably help this one. The assessment will become less important if all of these people are switched over to alternative therapies.

(Laughter.)

CO-CHAIR BASKIN: Okay. I appreciate that. But even those with added therapies, steroid sparing therapies, may still remain on steroids. They can't all get

off steroids, even with the other therapies. But let's not go there.

All right. Then I guess we're going to move on to our last measure. Now, we have a time issue here.

MS. WILBON: So, a couple things. It's time for lunch. Lunch is out, so we have a few options. We are about 15 or 20 minutes behind. We can have lunch, keep going.

CO-CHAIR BASKIN: I'll suggest that let's get 10 minutes to get lunch, bring it back to the table here. It's wraps, sandwich-type things, so there's no reason why, after 10 or 15 minutes, we couldn't start discussion while we're eating.

Okay?

MS. WILBON: Well, quickly, before we break, again, this is a pilot group. So we have actually an evaluation team that's working internally to try and help us gather some information about the process as you've experienced it.

And so I think some of my NQF colleagues are in here, Lisa and Helen, and they have a short survey they'd like you to fill out while you're working on lunch. It's five questions. It's really brief, shouldn't take much of your time.

So while you're eating and gathering your things, they're going to distribute the survey, and they'll collect it from you before we start discussions again.

CO-CHAIR BASKIN: So now that means 15 minutes before we start the conversation, because you get five minutes to complete the survey.

MR. AMIN: I also want to clarify that our conference center staff distribute a survey -- I hate to over-survey people, but they distributed a survey that's on your desk. This is a survey that they're handing out now. So are they going to do a little orientation to the survey?

MS. WILBON: No, it's just five

questions. It's pretty straightforward.

MR. AMIN: So the one that is being handed out now is the one that's -- not that any one is more important than another --

(Laughter.)

MR. AMIN: -- but that would be the one that we'd want you to focus on. Thank you. Unless you didn't like the food, and then feel free to fill out the other one as well.

(Whereupon, the meeting recessed for lunch at 12:13 p.m., and was resumed at 12:32 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

(12:32 p.m.)

CO-CHAIR BASKIN: Why don't you go ahead, then?

MEMBER BORDEIANOU: So, Measure C2065 is a little bit different from everything we spoke about here before, because it actually focuses on measuring, not individual providers, but hospitals, and it focuses on measuring an outcome as opposed to measuring a process.

And what they propose to do is to look at the number of in-hospital deaths, from hospital to hospital, caused by GI bleed. And they don't stratify in terms of what the cause of the GI bleed is -- is it variceal bleeding, or diverticular bleeding -- so it's a very generic outcome measure.

So as far as the impact -- we're still going through the same motions, right, when we discuss it? As far as the impact, GI bleeding is a very common problem, and if you

include every form of GI bleeding you're going to get a huge number of patients that are affected.

And yes, there is a mortality rate associated with GI bleeding. And on page 4, they discuss the rates of mortality, and they say that they haven't changed much in the last 14 years, but then they mention the health care cost and utilization project, and they say on that project they saw a decline in the rates of bleeding. So that sort of goes back to the second issue of the gap, which we can discuss later on. But from the standpoint of the impact, I would say that the impact is high.

OPERATOR: Excuse me. Dr. Romano has rejoined.

CO-CHAIR BASKIN: Dr. Romano?

DR. ROMANO: Yes, I am here. Thank you.

CO-CHAIR BASKIN: Hi. We were just getting started on this one, so let's

back up for a second and give you a few minutes to introduce this measure to us, please.

DR. ROMANO: Yes, certainly. It's a pleasure to be here. My name is Patrick Romano. I'm a general internist based at UC Davis in Sacramento, representing AHRQ today. This is a risk-adjusted outcome measure, a mortality measure, as you've heard. Perhaps the only outcome measure that's under discussion by this panel.

The focus of it is on in-patient mortality among patients who were admitted with both upper and lower gastrointestinal hemorrhage. It is one of a suite of similar risk-adjusted outcome measures for major conditions and procedures that are offered as part of the AHRQ quality indicators program, so there are similar NQF-endorsed measures for heart attack mortality, heart failure mortality, pneumonia, and stroke mortality, as well as a couple of procedures.

The basic approach here, as I think people understand, is that it uses administrative data of the type that hospitals collect for their own internal purposes, as well as for billing purposes and reporting for state health data agencies. The data that AHRQ actually uses for estimating and testing the indicators is the data that comes from state health data agencies, through what's known as the Health Care Cost and Utilization Project, so these are data that are widely used for research purposes, as well as generally tracking health system performance and clinical epidemiology of major conditions.

The risk adjustment approach is based on the 3M APR-DRG system, which incorporates a variety of factors related to the severity of the patient's condition, including comorbid illnesses, as well as manifestations of the particular type of bleeding, such as whether it's an esophageal, variceal bleed for example, or a lower GI

bleed.

That's built into the APR-DRG system, and basically 3M has an arrangement with AHRQ to make available limited licenses for free, so that architecture is available to those of you who are interested in the details.

So this measure has been available and in use for several years, and we're pleased to have this opportunity to discuss it with the National Quality Forum for possible endorsement.

CO-CHAIR BASKIN: Thank you very much. Is there any question for the developer or the measure steward before we open our discussion?

Zahid?

MEMBER BUTT: Patrick, in the section 1b.4, there is data that is stratified for disparity analysis. Do you know why the race/ethnicity was just observed rate and not risk adjusted in the data that's presented?

If I'm interpreting it correctly.

CO-CHAIR BASKIN: I mean, the disparities data that's asked for is just to help us determine whether there are disparities related to this concept or not. Whether it's part of the risk adjustment is an entirely different issue of whether that's an appropriate risk adjustor.

MEMBER BUTT: Yes, but I think that the question is whether the difference is on a risk-adjusted basis or not, so you can make a different conclusion based on that whether the disparity exists or not.

DR. ROMANO: Right. I'm not sure that I can answer that question right off the top of my head. I'm not sure if anyone from our analytic team is on the call.

I think the focus here was really, in terms of the gap analysis, on the disparities across different types of hospitals. And you'll not substantial differences between teaching hospitals and

non-teaching hospitals, between small hospitals and larger hospitals, and between rural and metropolitan and urban hospitals. So those analyses are all risk-adjusted.

CO-CHAIR BASKIN: Thank you very much. So we'll return, then, to our discussion. So, Liliana, we're going to talk about the impact?

MEMBER BORDEIANOU: The summary of the evidence with regards to the high impact is on page 4 if you guys want to see. And essentially, the discussion is about how the GI hemorrhage in general is a very common medical problem. Pretty much any hospital, small or large, encounters it. The mortality rate, depending on the diagnosis, could be as high as 10 percent. And so it is definitely a high-impact measure in my opinion.

CO-CHAIR BASKIN: Any questions or concerns regarding that?

(No response.)

CO-CHAIR BASKIN: Then I think

we're ready to quickly go to a vote regarding impact. This will be another one where there's four choices: high, moderate, low, insufficient evidence.

So, those that think this meets our requirement of high impact, raise your hands.

(Show of hands.)

CO-CHAIR BASKIN: That appears to be everybody, so that's 14, which means obviously zero moderate, zero low, and zero insufficient.

The evidence?

MEMBER BORDEIANOU: My understanding is that we don't discuss evidence, because it's not a process measure.

CO-CHAIR BASKIN: Yes, it's an outcome. So what do we do with the outcome, that's my question. That's why I turned to you, is what do we do with evidence here?

MR. AMIN: I apologize. So what we're looking for is --

DR. PACE: Basically our criteria for health outcomes is that you don't need to present the quantity, quality and consistency of the body of evidence, because outcomes are generally influenced by multiple processes. And so we asked them to provide a plausible rationale or connection between processes and structures to that outcome.

MR. AMIN: And that's on page 10. Sorry, it took me a second. That's on page 10, 1c.2.1, where we asked them to provide a rationale between a relationship between the health outcome and at least one structure, process, intervention or service, the goal here being that there is at least something that the health care community can do, or the measured entity can do, to influence this outcome. That's the rationale you have.

MEMBER BORDEIANOU: So on page 1, 2, 3 and 4 of the addendum that everybody has, the proposal sort of goes through a variety of different interventions that one could use to

improve mortality. And some of them are medical therapies, some of them are systems therapies, and a lot of them are randomized controlled trials that are being quoted.

So there are definitely a variety of different interventions that a hospital could implement, if they are not implementing them already, to control and prevent mortality. I think that there is plenty of evidence.

CO-CHAIR BASKIN: So, plenty of evidence to support that there are interventions that will result in improved outcomes.

Any comments or questions regarding that?

MEMBER BUTT: I think in this, the highest evidence is in that subset of esophageal/variceal bleed and massive lower GI bleed. So I guess the question will be, would it be helpful to stratify this measure along those lines? Because it dilutes out the

impact when the entire --

MEMBER BORDEIANOU: I was saving my comments about the stratification and what the model is going to be, and how they're going to account for various comorbidities and other factors. I think that how this will be measured is key. As we're going through the motions, I figured that would be Phase 2. But is it high impact? Yes. Are there things that one could do to improve outcomes? Yes.

CO-CHAIR BASKIN: I think, though, part of your point, Zahid, if I'm reading it correctly, is that some of these interventions are obviously more impactful in terms of health outcomes than others, and those interventions are specific to certain diagnoses within this large range of diagnoses.

And I guess the question is, do these interventions and outcomes appreciably affect the entire measure as opposed to just small snippets of the measure, small snippets

of diagnoses within the measure?

MEMBER BUTT: Right. And they mention that in the submission itself.

CO-CHAIR BASKIN: So are you comfortable that there are interventions here that substantially affect the measure as it stands?

MEMBER BUTT: I think so. I think that it might actually be helpful if there is -- and perhaps that will come out later in the comment section. But I think overall, there are interventions that do improve the outcome.

CO-CHAIR BASKIN: Okay. So we're voting yes or no, the evidence meets what we need?

DR. PACE: Right, that it meets our criteria, which is basically for a health outcome that there's a link to at least one health care service or treatment or intervention.

CO-CHAIR BASKIN: Right. That's what I said. Thanks. I appreciate your

saying it for me. I couldn't have said it nearly as well.

So, everyone's comfortable with the question, so that we can vote?

(No response.)

CO-CHAIR BASKIN: It doesn't seem like any heads are shaking in the wrong direction. Okay.

So yes would be that the evidence meets our criteria, and our criteria are a little bit difference in that the evidence shows a link between interventions and health outcomes. 2 means that there's inadequate evidence, and 3 means that it's inadequate evidence but that we think that that evidence does exist.

So, let's vote. All those voting 1, that the evidence has been submitted and exists?

(Show of hands.)

CO-CHAIR BASKIN: Okay. That's unanimous, so that means, obviously, there are

zero votes for the other two options.

And then we move on to performance gap. So Liliana, thank you.

MEMBER BORDEIANOU: So on page 5 and 6 of the proposals, we get a breakdown that suggests that the odds ratio of bleeding ranges anywhere from 17 to 22 based on the type of the hospital, and then there is a gap based on age and social/income, and Medicare versus other insurance. That's page 6. And again, the odds ratios are anywhere from 14 to 25 in the uninsured.

So there is clearly a gap. The concern I have is a more generic concern, and that's, I think, where Zahid is getting at, is that we don't know how they're adjusting for this. I think we need to see the formula on how this is being calculated.

CO-CHAIR BASKIN: Yes. Actually, I think that probably goes in the Stage 2 discussion, to discuss the risk adjustment, whether it's appropriate, and whether it's

actually measuring what it says it's going to measure. So I don't really think we want to start down that path, because we can't finish that conversation here today.

But I think that, in terms of the gap, it's not just that the gaps were presented, but they're statistically significant. I mean, because a range of 17 to 22 may or may not be significant, but apparently it is here.

Any particular comments around the gap?

MEMBER BUTT: I just wanted to clarify what I was saying. There is actually very good documentation and very solid risk adjustment methodology that they use. So that's not the issue. The question I was asking was that, in the gap section, where they have basically -- in the disparities section, they break down a whole bunch of different categories where they stratify the results. All of them are risk-adjusted except

for the race and ethnicity breakdown.

And my question was, is there any specific reason -- because there seemed to a difference between white, black, hispanic, asian, on the unadjusted rates. But whether that would hold up when the risk-adjusted -- because the measure is risk-adjusted, and they risk-adjust everything else. And my question was, is there a specific reason why that was not risk-adjusted?

But there is definite scientific validity, and as a matter of fact these IQI and PSI measures are very well-thought-out and done. Extensive documentation is used for the risk-adjustment methodology that they use.

DR. PACE: I think that's a good question, and we can certainly ask that the developer make that clear when this comes back, unless it's something that will really hold you up.

MR. AMIN: Is the concern here also that race might be in the risk-adjustment

model, that that's the reason that they reported it?

MEMBER BUTT: No. Sometimes the observed rate can actually change when you risk-adjust it, and that's the reason to risk-adjust a rate, because based on comorbidities, and whatever other -- and they use the APR-DRG classification system to risk-adjust based on that.

In other words, so right now, the unadjusted rate -- if I, again, interpret this correctly, the mortality rate for blacks is .09 and whites is .14. But once you risk-adjust it, it might be different.

CO-CHAIR BASKIN: John?

DR. ROMANO: I can actually address that question now.

CO-CHAIR BASKIN: All right. Thank you.

DR. ROMANO: It was really just a fluke, to be honest. Race and ethnicity are not in the risk-adjustment model. We do

adjust for age and gender, as well as the transfer status of the patient, whether the patient was transferred in from another emergency room or hospital.

But we do not adjust for race/ethnicity, and so that requires a separate stratified analysis, and we just ran out of time to do that before the submission document went in.

CO-CHAIR BASKIN: John?

MEMBER MORTON: I was going to say that, oftentimes, risk adjustment is not made because these are administrative databases, and race is missing quite often in those databases. Nationwide inpatient sample, it's missing upwards of 20 to 30 percent depending on which one you're looking at. So that might be one reason. But we already heard from Dr. Romano.

CO-CHAIR BASKIN: Judith?

MEMBER TOBIN: Just a question, because this comes up a lot with CMS. If you

risk-adjust for things like race and ethnicity, then you're risk-adjusting away potential disparities.

CO-CHAIR BASKIN: As well as socioeconomic risk adjustment. This has come up in the CSAC on many an occasion, and the tendency has been to stay away from risk-adjusting based on those, because it does hide those disparities and potentially hinders improvement in those situations. Oh, and it's in the NQF guidance as well.

So yes, I think this was more of interest, not that it would stop this measure or the appropriateness of this measure, but since the data apparently could be available, it was of interest to us, as Zahid said, are there differences if you had risk-adjusted it? Not risk-adjusted the actual measure result, but are there actually -- when you're looking at it, just for informational purposes, is there an issue based on race or ethnicity?

MEMBER BUTT: Not to get too much

into the weeds of this and prolonging this any further, but I was actually looking at this more as a stratification of risk-adjusted rate, rather than using these components to risk-adjust itself. Again, I was looking at this as more of a stratification of risk-adjusted rate that I thought was being done using APR-DRG and a couple of other things that he mentioned.

CO-CHAIR BASKIN: All right. I think we're ready to take a vote here on the performance gap. We have four options here: high, moderate, low, insufficient.

So, all those raise their hands who feel that performance gap was demonstrated high, considerable variation.

(Show of hands.)

CO-CHAIR BASKIN: Ten. How many think moderate?

(Show of hands.)

CO-CHAIR BASKIN: Four. And that means zero lows and zero insufficients. So

then I think we can move on to whether we recommend the approval of this concept or not.

I'm looking around to see if anybody needs to make a comment. Otherwise, we can just take this directly to a vote. I think we'll go to a vote, then.

Those in favor of approving this, vote yes. Raise your hands.

(Show of hands.)

CO-CHAIR BASKIN: It appears to be 14. It appears to be unanimous. So that would be zero nos. Thank you very much for that.

Any comments back to the measure steward that anyone wants to make at this time?

MEMBER BORDEIANOU: My only comment is the numerator and the denominator, they are only including patients that have as the first diagnosis GI bleeding. There is a lot of room for manipulation of that.

If the hospitals learn that

they're being measured on GI bleeding, they can code the first diagnosis as myocardial infarction, or whatever the cause of death was, as opposed to what the presentation cause was. So I think that this needs to be heavily considered in the feasibility part of the discussion.

CO-CHAIR BASKIN: Thank you, and I think that's come up with similar comments with the other similar measures that have the same issue.

Zahid, comment?

MEMBER BUTT: So I would strongly encourage them to look at if they could stratify it by variceal bleeding, because I think that has, probably, as a subgroup, the biggest impact.

CO-CHAIR BASKIN: So essentially you're suggesting that there may be large subgroups here for which there hopefully could be enough of a denominator that it may be of some interest to --

MEMBER BUTT: Or even if they could somehow stratify this.

CO-CHAIR BASKIN: -- to stratify out the components of the measure based on diagnosis.

MEMBER BUTT: Yes.

CO-CHAIR BASKIN: Any other comments for the developers?

(No response.)

CO-CHAIR BASKIN: Okay. I would think, though, that anyone that's actually implementing the measure could do that stratification, really, themselves, if they wanted to. Although the stratification could itself be a measure, of course, but you would have the ability to do that. If you have the ability to perform the measure, you have the ability to do that as well.

Okay. Well, thank you very much. That concludes our review of the measures, and right on time. We've already had our lunch. We're going to open it up for member comment

and public comment. Member comment are those in the room, so if anyone in the room would like to make a comment, you have a microphone. Just state who you are when you make your comment, please. Thank you.

DR. PARK: Walter Park again, on behalf of the American Society of Gastrointestinal Endoscopy. We just wanted to make some brief verbal comments regarding measure concept 0259 by AGA. On behalf of the ASG, we do support the passing of this concept.

We do share some of the concerns raised by some of the members regarding further clearance on the denominator, and as we look forward to Stage 2 we only request or look forward to seeing the developer define the concepts in a manner that is registry-neutral. That would allow our fellow gastroenterologists who do not participate in the AGA registry to be able to comply with this measure.

CO-CHAIR BASKIN: Thank you. Any additional comments from anyone in the room?

(No response.)

CO-CHAIR BASKIN: Then we can open it up for public comment on the phone. Is the line open, operator?

OPERATOR: Yes, sir. All lines are open.

CO-CHAIR BASKIN: Thank you. Any comments?

(No response.)

CO-CHAIR BASKIN: Hearing none, then we'll close the public comment portion and we'll move on to the next topic area, which is potential for harmonization and identification of gaps in the GI measurement. Taroon, did you want to lead this or get us started?

MR. AMIN: Yes, I will just pose a few questions, actually, to the group. I think the two -- and Ashlie, please jump in here if there's anything else that you want to

add. It sounds like there were two related concepts from today, 0658 and 0659. 0659 was not recommended to move forward.

No, I have that wrong, sorry. It's C2059 and 2062. I apologize. So the question I have is, while 62 didn't move forward, can we get some clarification on exactly what the recommendation would be? I know we mentioned that we wanted to have components of 62 incorporated into 59, so maybe we could just have a little bit of discussion on that, of what you would like to see there, if anything. And then we go back to our discussion that we began yesterday, which is on 0653, the chronic liver disease with the hepatitis A vaccination, and look at -- Ashlie, did you have something?

MS. WILBON: It was 0635, instead of 53.

MR. AMIN: I apologize. I'm messing up all these numbers right now. So it's the chronic liver disease/hepatitis A

vaccination, along with the patients with hepatitis C who get hepatitis A vaccination. We put the side-by-side table up on the screen for you, to be able to see the side-by-side. The one that's the patients with hepatitis C who had hepatitis A vaccination, that measure was not in this project, and it's actually being reviewed in the ID project next door, but they are related.

So the question here is, is there anything that you would like to see related to how these measures relate to one another prior to moving into Stage 2?

CO-CHAIR BASKIN: So, let's talk about that combination first. Those two measures, I mean. I don't mean combining them, necessarily, but those two measures with the hepatitis A vaccination.

So not knowing the other measure off the top of my head, the one that's already in existence regarding hepatitis C patients getting hepatitis A vaccination, the

denominators, can we look and see, are they reasonably the same -- is one denominator essentially a subset of the other measure, the chronic liver disease measure, and completely included in it?

So all patients with a diagnosis of hepatitis C, and the other is patients diagnosed with chronic liver disease. And I believe the chronic liver disease group in relation to hepatitis C -- how do they get into that denominator?

So, basically, if you've had chronic hepatitis C -- and is that really the -- I don't know the interpretation here. Is this really the same populations?

MEMBER BUTT: So it looks like, when I looked at them side-by-side, because I think I did the one yesterday for the new measure, there are some significant differences. The big differences, at least from what I can see in this side-by-side comparison, is, number one, the data source is

a big difference. The new measure includes patient-reported survey data, all sorts of EHR data, HIE data, whereas this one is pretty limited to the type of data that would be in possession of a practice, which kind of leads into the next key question, which I think was raised yesterday as well: what level is this applicable at?

The new measure is being applied at the population level, whereas this old measure is being applied at a clinician, individual physician, level. And the importance of that is because that's where that sort of denominator comes in, that for the physician level you have to actually attribute it to a physician, and typically it's done through CPT, office visit type of data, that if you've had two office visits or one office visit, that sort of gets counted in the denominator.

But here in this other one, the denominator is sort of a big, large sort of

multiple data sources -- patients could have just self-reported, et cetera, et cetera. So the denominators are really very different in these two, and the level of application, at least the way it's presented, is totally different.

So, those are some of the key differences to me that would present, I think, some problem in harmonizing these. Also, the old measure is only for hepatitis C, whereas this new measure is for all chronic liver disease. So that's another big difference.

DR. PACE: So I think one of the questions is, what does the evidence say? Who should be receiving the vaccination? Should it just be restricted to patients with hepatitis C, or is it all chronic?

MEMBER BUTT: As we saw yesterday, the evidence would suggest -- and all three guidelines were very consistent -- that it should be for all chronic liver disease. So I think that the evidence would suggest that

it should be for all patients. So the hepatitis C would be just a subset of that.

CO-CHAIR BASKIN: But if you did apply attribution logic to those in the chronic liver disease measure -- because if a patient had seen a physician for two days you could do that measure at a physician level.

MEMBER BUTT: Right. So one harmonization might be that the old measure could expand its denominator to include all liver disease, and that would actually accomplish that goal. I don't see why it couldn't, because the body of evidence is there.

The new measure, obviously, is being applied for a different reason from, at least, what is being presented. So that would have to be evaluated, whether it actually does represent as a population measure or not, in Stage 2. But certainly, I think, in terms of trying to accomplish a part of what it was trying to do, would be to include all chronic

liver disease in the existing measure.

It's otherwise really well done. The existing measure seems to have been well thought-out and well done.

CO-CHAIR BASKIN: Any other comments?

MS. WILBON: Zahid, can you clarify what you mean by existing measure? Because they're actually both maintenance measures, so I was a little confused.

CO-CHAIR BASKIN: The measure we reviewed yesterday was not a new measure. It was a maintenance measure. So he was describing that was the new one, because it was the newest for us to discuss.

MR. AMIN: Just to clarify, when he was referring to the new measure, he was referring to 0635.

MEMBER BUTT: Yes, 0635 is what I was referring to as a new concept.

CO-CHAIR BASKIN: We don't normally have this back-and-forth, but I know

you have something to say that would probably be very relevant to this, so please just introduce yourself and go ahead and speak to that. Thank you.

DR. ANTMAN: Thanks. Mark Antman for the AMA-PCPI. Just to note that, knowing that this discussion would come up today, we did discuss the idea of potentially harmonizing with 0635 with our hepatitis C workgroup cochairs, and they certainly agreed that it would be appropriate to -- that hepatitis A vaccination obviously is supported by the evidence for all chronic liver disease. So we are interested in the recommendations of this committee as to how we can harmonize with the active health measure, recognizing that there are, as Dr. Butt pointed out, some data source challenges.

CO-CHAIR BASKIN: Okay. And I think that this connection can be made outside of this meeting, off-line, and see if those discussions can occur. Thank you, though, for

that comment.

Any other comment in the room here regarding these particular two measures, this pair?

(No response.)

CO-CHAIR BASKIN: Okay. So there seems to be some opportunity here that can be explored. The other pair that you mentioned was the IBD. So one that made it through today, and one did not, but there's a question of the fact that the IBD measures regarding the cortico-sparing therapy and those that are on chronic corticosteroid therapy, regarding an assessment for bone loss, is there an opportunity to incorporate the assessment or something to do to address the issue of potential bone loss and potential treatment or not? Only because if you're looking at the same populations in the denominator, and is there a way to do that?

Now, one of the issues is that just adding it in doesn't work, because we've

decided that that measure doesn't meet our criteria. So unless there was some way to get at more of what we were considering the outcome or the treatment, or something other than just performing or not performing an assessment for which there would be an evidence base to support that -- it's certainly a similar population, and would make some sense to enhance the other measure.

But I'll open that up for anyone that has anything more to say about that than I've just commented on.

MEMBER BUTT: I just think it would be hard to combine the two, because they're really -- the steroid sparing and the use of immunomodulator or anti-TNF therapy is sort of a different objective there, and this probably would be difficult to fit into that, is the way I think about it.

CO-CHAIR BASKIN: About the only thing they have in common is the same denominator and the fact that they're both

appropriate steps to take, but very different steps in terms of what their goals are. So I understand where you think that maybe it doesn't make sense to consider a combination there.

Any other comments?

MEMBER BORDEIANOU: Unless the measure in general is reformatted as a discussion about the risks of long-standing steroids, and a discussion about treatment options such as steroid sparing therapies, measurement of complications, et cetera, surgery.

CO-CHAIR BASKIN: So essentially the comments we've made regarding as to what we would have liked to have seen for that measure to even come back as its own measure is still valid, whether it be combined with another measure or not combined with another measure. Right.

No further comment on that?

(No response.)

CO-CHAIR BASKIN: Then a gaps in GI measurement discussion. Do you have anything in particular, other than we're asking?

MR. AMIN: Yes, exactly.

CO-CHAIR BASKIN: I don't have a list of the subset of other GI measures that exist, other than the ones we've discussed.

MS. WILBON: So there's kind of a similar scenario with the two endoscopy measures that were both submitted by AMA-PCPI, 0658 and 0659. We have some comparison tables we can hand out, but I believe -- I think one of them was approved and one of them was not. Let me just double check here for one second.

Oh, no. They were both approved. So I guess the question is just to kind of bring it to your attention that they are both focused on polyp surveillance, and whether or not there's any room or discussion about whether or not harmonization can occur between those.

CO-CHAIR BASKIN: Certainly they're both talking about colonoscopy and interval under different circumstances, and one could argue that there's the possibility of having a measure out there that says "Hey, of any colonoscopy that was performed, depending on the results, was the appropriate interval either suggested or occurred?" And I guess that would be one way to do that. I'm not so sure that it's practical at this point to do that, so I'll just point that out as my own personal view. But it would be neat to have a measure that essentially included every colonoscopy, and was the appropriate interval adhered to or not adhered to, and that would be a great measure. But I'm sure there'd be some tremendous implementation and practicality issues on whether that measure could actually be performed and be accurate. But that is an ultimate, I think, fairly decent composite measure, to be honest with you.

MEMBER SCHOENFELD: In my opinion, it's not practical for harmonization at this time. And what I would keep in mind is, remember that the -- 0657, Taroon? Which is the colonoscopy screening one.

MR. AMIN: 0658.

MEMBER SCHOENFELD: 0658 refers to making a recommendation by the endoscopist. After you do a colonoscopy for somebody who's had a normal screen, you make a recommendation to say it should be done in 10 years. 0659 states that, if I'm doing a colonoscopy because a person has a history of polyps, that I am documenting at the time I do the colonoscopy that it's been at least three years.

We're really talking about two very different aspects of minimizing overuse of colonoscopy. So again, in my impression at this time, probably not appropriate for harmonization.

CO-CHAIR BASKIN: Any other

comments to be made regarding that?

(No response.)

CO-CHAIR BASKIN: I mean, I'm all in agreement about the practicality of doing it, although I have to admit, at the time, I suggested that the other measure be reversed, and I still think that that's a better way to go. So I'm going to say it again, because I've got the microphone.

MEMBER SCHOENFELD: And maybe AMA-PCPI can take that one up, because I am in agreement with you about that as a general theme.

CO-CHAIR BASKIN: Okay. And with no other comment, then I think -- once again, if there's anyone that has a recommendation or a suggestion or a request, or an identification of gaps in measurement that they think would be reasonable for a developer to fill, now's a great time. But any time is a great time for that. That information can be relayed back to NQF. They are always

requesting any feedback regarding gaps in measurement.

So not seeing anyone quick to raise their card to be able to identify a gap, then I think we'll move on. We want to get some pilot feedback at this point.

MR. AMIN: Ashlie and I will tag team on this question, and Karen's here as well. But as we described at the beginning of yesterday -- I'll actually take a deep breath here. We're done with a lot of the heavy lifting.

MS. WILBON: Good job, by the way.

MR. AMIN: Thank you for all that.

MS. WILBON: Way to push through.

MR. AMIN: I know this is a tall order. So this is more of a reflection period on kind of where we've been over the last two days, and kind of hearing your feedback on how this pilot has been working. And in an overall standpoint of how this process is different than the current NQF process, since

many of you are new to the CDP process, is that we implemented a number of different components.

The first -- and we had some sidebar conversations about it as well -- is a technical review period which happened prior to measure submission, where we asked measure developers to submit at least one concept to Karen Pace and Alexis Forman, who did a thorough review of the evidence and a number of different components of the measure, and provided technical feedback on areas that needed to be expanded upon or needed more clarification.

We also split the process in two, which is why we call it a two-stage process, in which we broke out the importance criteria away from scientific acceptability, usability, and feasibility, which you'll evaluate in your second stage. And there are a number of other tools that we developed to support this process.

So what we wanted to do now is to try to get some feedback from you -- again, knowing that many of you may not have participated in the typical CDP process -- on a number of key questions related to some of the changes that we implemented.

And I guess one of the first questions that I will start with is that we made some assumptions on how we can actually define a concept, and the way we defined a concept was around the numerator, the denominator, exclusions, usability information, taxonomy.

Was that enough to really get a sense of what the measure concept was that you were trying to evaluate? Was that a sufficient amount of information to evaluate the concept? And was there information that you didn't review, or you did not think it was necessary for us to collect from the developers?

I think one question that seemed

to occur, one piece of feedback that already we seem to have gotten, is that while we looked at the information that was presented, it gave us a sense of the concept but we didn't actually evaluate, necessarily, the construction of the concept. So particularly the numerator, denominator, or if there are multiple components in the numerator and denominator, that was not explicitly evaluated in this process, because we're just looking at those importance criteria.

So I guess I'll start -- do you want to do each of them individually, or should I go through all the questions? What do you think?

MS. WILBON: Let's do them individually.

MR. AMIN: Yes. Let's stop there. And just so you know, I have three slides with sets of questions. Just so you're not overwhelmed.

CO-CHAIR BASKIN: So, comments?

And I'd like to make an initial comment, in that one thing that I struggled with a little bit and would have liked submitted was a brief statement by the developer on essentially what the intent was. I mean, what did they expect that this measure would do, or how did they expect that this measure would result in improved health outcomes? It wasn't always so clear to me.

And in fact, if we had asked the developer to say what, then maybe somebody that came in and said "Well, doing an assessment somehow or other" -- I mean, force them to -- I don't mean force in a bad way, but because we don't get a lot of back and forth discussion with the developers at this stage, to kind of get a feel for "Did you really think that this, somehow or another, links to some change in performance, or that people's behavior is going to change, either patients or doctors, whoever it is that you're measuring here?"

So even if it's just a paragraph, three or four sentences, with a kind of intent and an expectation of how you think this measure would play out in terms of affecting care, that would have been helpful to me.

MEMBER MERGUERIAN: Just to add to that, I would totally agree. I think linking the measure to outcome measures, and actually having the developer think about what types of outcome measures they're trying to -- or they will develop in the future. The other thing is really looking at that from a patient perspective, looking at the value of this measure as far as a patient is concerned. You know, patient satisfaction, other types of measures -- you know, value, patient values. Because none of these concepts really looked at it from a patient perspective.

CO-CHAIR BASKIN: I lost track completely. So I think, John, yours was up before.

MEMBER MORTON: I was going to

make the same point about the patient preferences, and making sure they're included. The other thing that came up, we didn't have a ready answer for, is to what degree does cost enter into any of this? And maybe it would be something good, to figure out what the playbook looks like, in figuring out what role cost should play. Maybe a bigger concept is value, cost and quality combined. So, just a thought.

MEMBER MERGUERIAN: There is actually a compass called the Value Compass that actually looks at four areas of measure: functional, satisfaction, cost. And so that's one area that I think -- it's developed by IHI, and you can actually get that. It's called a Value Compass.

CO-CHAIR BASKIN: Well, we have to get a feel for whether cost effectiveness plays a part in our decision making at all or not, to be honest with you. Because it's not something we asked for data on, and it's not

something that's intuitive, unless you're an expert in that particular activity.

MEMBER REYNOLDS: Right. So one of the things that you had asked for, and you had supplied, is usability info, but that was often not completed on the forms. And that would be a chance, so the people could put down what it's going to be used for, how it was going to be used, and it wasn't clear to me that we had a chance to really discuss that or evaluate that now. Now, granted, it might be part of the second stage.

And then the other part that people talked a little bit about is this concept of the proximity to the outcome. It came up a couple times, and I just wonder if pushing that to the first part of the two-step process would also be helpful. Because again, it might be high-impact and whatnot, but if it's really proximal to the outcome, we might want to flesh that out ahead of time, and not go forth on the second part.

CO-CHAIR BASKIN: So essentially asking that, if your measure is not proximal to the outcome, why isn't it proximal to the outcome.

MEMBER REYNOLDS: Well, the reasons --

CO-CHAIR BASKIN: The reason your measure had to be so distal.

MEMBER REYNOLDS: There's no real point in our evaluation to address that. I mean, we talk about that at the end, after we sort of voted it through. Like "Gee, this would be better if we were looking at, in fact, the number of colonoscopies that were done, rather than" et cetera. I think that there's a point where we could address that further before we get to the next step.

MEMBER BORDEIANOU: Maybe I'm saying the same thing in a different way, but any medical problem that we'll be discussing here is going to be high-impact. It seems like that's a no-brainer a majority of the

times. But what is not always being clearly discussed -- or maybe I'm missing it -- is "Does the particular measure have a high impact when performed?" I.E., doing a physical exam before surgery changes the outcome. This is where we really need to be digging into more.

CO-CHAIR BASKIN: That's essentially what the evidence review is all about, except to say that in many cases, for things like physical exams and asking a particular question before surgery, or before a procedure, there's often a little lack of evidence. And I think that's a problem, and we saw that over and over again. We struggled with that.

MR. AMIN: Right.

MEMBER MORTON: The other thing I was going to make, for the continuing measures, I know it's something that we ask for, but it would be great to have more emphasis on what has happened since that

measure came into play. I think that's key, closing the loop. We think of all these measures a little bit in isolation, but what happens in real practice? I think that's really important, particularly as we see some of the older measures become pretty mature. And it may be time to sunset some of these, or it may be time to say "You know what? That was good four years ago, but it's not good now."

CO-CHAIR BASKIN: When is the discussion of whether a measure should be -- the discussion should be whether it should go into reserve or not. Is that really a Stage 1 discussion, or is it a Stage 2 discussion? It's kind of hard to even discuss the measure. If we don't discuss it in Stage 1, why are we even talking about it for maintenance in some cases?

MR. AMIN: One of the other components that we're testing here is to take out the evidence form as the attachment, to

make it a little more clear than we've had it in the past. Just some experience in reviewing the evidence form: was it clear what type of information that we requested and why? Did you feel that the format of the evidence form was conducive to completing your reviews in an organized fashion?

So it's broadly about the evidence information that we asked for. Was it clear to you? And was it clear to evaluate, just in that sense?

MEMBER REYNOLDS: I think it was a little bit unclear. I think it was unclear to the developers exactly what we needed and what they needed to supply. And we struggled with that. I also think that when we are evaluating it, it could have been a little bit more helpful if we had stuck a little bit more closely to the quantity, quality and consistency, which when we did the preliminary evaluations, we were sort of asked to specifically rate those individually and then

a global thing.

I think that if we had had that opportunity, we would have been a little bit more strict on the evidence going forth. So if we had specifically had to say "Can we see that there are four or more studies in the form? Were they consistent? Blah, blah, blah," you'd at least get a little more granularity on what the issue was, rather than this global "We think there's enough evidence of pretty high quality."

We might have eliminated a lot more measures if we'd been a little bit more strict.

MEMBER TOBIN: I would second that. Just as a non-voting person, but observation, there seemed to be a lot of deviation from the strict rules that were established for the quantity. And the other was, is there compelling evidence, if it doesn't meet the criteria? And I found that a little -- because you're not obligated to

present evidence, but you could say that you think there's a compelling reason to still push the measure forward. I found that confusing.

DR. PACE: Can I ask a question? And this is actually broader implications than the pilot, but what would you think if NQF just took the hard line of "We're not accepting measures that are based on expert opinion and consensus, and that we want measures focused, that are proximal to the outcome, with that evidence-outcome link?"

You know, the reason for all that language is the continued push-back of wanting these more distal process measures. So this would go to a higher authority, but since it's brought up, I'll just see what your thoughts are.

CO-CHAIR BASKIN: Go ahead, John.

MEMBER MORTON: I agree that when you grade evidence around expert opinion, it's never very high. But there are going to be

occasions where there's compelling reasons to have expert opinion, because there's simply no data yet and the need is high to have some sort of quality measure out there. So when there's gaps like that, I would be reluctant to exclude it altogether. I think we have to grade it and get a better idea of "Does this rise to the occasion when we accept only expert?"

And there may be circumstances where there's a real compelling quality need that we only have expert opinion. But I agree it's not the best, but I wouldn't do away with it altogether.

CO-CHAIR BASKIN: I am just going to start at the end and work up in order since I didn't watch you put up the cards.

So go ahead, Zahid.

MEMBER BUTT: Yes, I agree with John, that probably it would be worth keeping it in. But I do also agree with what Judith was saying, that I think where it may be a

sort of opportunity to make the body of evidence section specific, where you actually have a small table that they have to fill out, the developer. Because some do count the studies in that section. Others don't.

So if you force them, or it becomes a requirement of filling that section, that they have to count the number of studies, they have to grade the quality, and they have to grade the consistency. So at least, whatever they present, they should do their part of it.

And then in the area of guidelines, that's where, to me, I had the most difficulty. Because one assumes, often, in practice, that practice guidelines, especially out of your professional societies, are the standard of care. Because we reference them all the time. You know, we do colonoscopies, and the first thing we say is "ACG, or ASGE, guideline says that I should do this."

So there seems to be somewhat of a disconnect in the perception of a guideline and, perhaps, what NQF is looking for here. And I don't know how that gets reconciled and harmonized, but at least in the short run, where there is a guideline, then the measure developer should provide all that information that backs up the guideline, so that at least you can make a judgment "Okay, this guideline is based on this number of studies and this number of randomized controlled trials," or "this amount of expert opinion."

I think if all of that is nice and concise and well laid-out in that 1c section, then it would make the job for the steering committee easier. Then you just have to sort of validate what is there, to the extent that you can. So I think that might be one other area.

The last comment that I'll make is that there seems to be a lot of duplication in the data that's presented. Like, there is

evidence for high impact, and then back down in 1c there is evidence again. And some of the developers are just repeating the same thing up there. They just reference the study, rather than take out the portion of the study that addresses high impact, the portion of the study that may only address a different section of the body of evidence.

So I think some of those things, if there could be some design of the form that sort of guides them through that process and makes it more clear as to what they have to provide, it would make the job of the steering committee easier.

MEMBER KOCH: So, to follow up on that, I think what the expectation of the developers should be is that they actually rate and grade the evidence. I mean, it's in their section, but the majority of the proposals didn't actually include that. Now, they could have, just like we ended up doing. They should have been expected to do that, and

they shouldn't get to submit something without that.

The issue in terms of the guideline, I think, as a process to this thing, I would suggest that things that are based on guidelines or not enough evidence should be considered later in the day or in a separate category. I think part of what happened to us is that the very first thing we did was spend 45 minutes -- and I'm not sure that that measure, if it had been presented later, with all the discussion we had, would have qualified.

So setting the day up so, if you have a brand new group, make the first one "This is a slam dunk, this is our best proposal, it's got great data." Then things that are coming back up for reevaluation, especially if they're -- you know, something that's just guideline-based should have a little asterisk. Three years later, the bar should be way higher and we should be seeing

way more data in order to substantiate that.

MEMBER MERGUERIAN: Again, I would agree. I would not take a hard-line approach, especially if you're going to delve into pediatrics, because there's really not a lot of data in pediatrics.

The second issue is grading the evidence, really having the developers grade the evidence, but then also giving them guidelines. Because there are two or three different grading systems, and so really just sticking to one grading system that you would then agree upon.

MEMBER LIGHTDALE: I actually agree with pretty much everything that's been said. My thought about consensus-based guidelines is that, right now, if we decided there wasn't enough evidence, we stopped and didn't ask about performance gaps. And I think it's okay to have a quality metric on something that there is basic consensus that it should happen, but there's also very good

evidence that it's not happening uniformly. And so stopping and not asking if there's a performance gap, I think, sort of defeated our purpose there. I don't know that the metric that we did that on actually had the performance gap evidence, but that could be compelling.

And then, also, with guidelines themselves -- of course, the corticosteroids and bone loss one was the one I was really looking at, but we've all been involved in guideline development, and that was a 2006 guideline from the AGA. Over the past six years, the rigor with which guidelines are being developed -- I think the understanding of the responsibility that the societies are taking on now has tremendously developed. And so a 2006 guideline was being held to a very different standard than a 2012 one, and really keeping an eye on that is going to be important.

MEMBER MORTON: I was just going

to add my voice to the chorus's, in that I think the idea is terrific about the summary table. There's clear criteria that's laid out, so why not have them put it out there? And it gives them a better understanding as to what we do, and it makes it, frankly, a lot easier to just ratify what's been done. So I think that's a terrific idea. The other great idea is "What has happened since," if it was expert panel. So I totally agree with both of those.

MEMBER BORDEIANOU: I'll just echo the feeling in the room, that we shouldn't say "No guidelines, ever." Because at least in surgical research, you will never have a randomized controlled study for a lot of what we do, and so if expert opinion will not count at all, you'll never have a measure of quality for surgeons.

MEMBER TOBIN: And I guess I don't want to give the impression that I think the criteria should be so rigid that if there is

compelling expert opinion, that that should be ignored. I think what I was weighing back and forth is, I'm not sure during the last few days if it was always applied evenly. And I think if I were a measure developer who had my measure rejected, I might think "Well, gee, had I had somebody else at the table, they could have made a really compelling argument." So it was just sort of this back and forth in my head, that I was sort of on the side of, if you were rejected, what would your response be?

CO-CHAIR BASKIN: Thank you.

MR. AMIN: And the last set of questions that we have is more on the preparation. The overall theme here is the preparation that staff were able to give you as steering committee members.

So the first question, given the project timeline, this is slightly tighter than our general CDP in terms of how much time you had to review measures. But the amount of

project timeline, the volume of information that we asked you to review, is there any suggestions that you have in terms of how we can better disseminate this information to you, in terms of format?

The webinars, I know many of you had difficulty with the Sharepoint site. Specific parts of the criteria that you found particularly difficult to understand? Is there any better information that we could distribute to committee members in genera?

MEMBER PELLETIER-CAMERON: So, most of the measures that were distributed to us were 12, 13 pages, which I think is reasonable. I mean, that's a volume that you can reasonably make your way through, given the number.

I just felt that some of them, although they were 12, 13 pages long, there were some that had full pages where there was no information filled out, and I almost felt bad for the developers that they didn't have

a chance to maybe utilize some of that space to their advantage, whereas there were other measures that were, again, a hundred and some pages long -- I'm not sure how that fits in there.

But I think keeping it a reasonable length is good, but maybe finding a better way to utilize the space so that there's not so much blank. And maybe that's just that they didn't bother to fill it out, but I think that there'd be -- keeping it the same length is good, but allowing them to utilize it better so that there's more there for us to read. Because more information's better, but without being excessive.

MEMBER SCHOENFELD: I mean, this is a more general comment, which is that I'm not totally sure why we combined GI with GU, except to the extent that I understand that you probably didn't have enough separate GI proposals and separate GU proposals to justify, say, doing this meeting -- I'm

assuming -- in terms of maybe having the CMS representative non-voting, the consumer representative, et cetera.

Having said that, with the way the proposals were distributed, to a large extent I'm not sure I contributed a whole lot as a person who focuses on quality improvement in colorectal cancer screening to all the GU discussions. And I'll let my GU colleagues comment on how much they felt they contributed.

And for somebody like, say, Mr. Ellis? Sure, have him both days. Have the GU people here on Monday and the GI people here on Tuesday. If the issue is a quorum, because you need a certain number of votes, I think we're being a little bit artificial here, to the extent that yeah, maybe I'm a vote in terms of discussing a GU proposal, but I don't necessarily think it's a very informed vote.

MEMBER GILL: So I think reviewing all these -- it was a lot of work to review

them, but I think what was perhaps even more taxing for the first time reviewer was trying to figure out the process. And I don't know if it's possible, or maybe there's confidentiality against it, but actually providing a whole measure to see how it flows, so we could just look at it, instead of having to figure it all out for each step, might have been easier for me, at least.

MEMBER MARKLAND: I would just like to add on one point. I agree in some ways with separation of the GI/GU, but I'd like to see if there's some primary care-focused measures, maybe have some primary care impact that has cross-cutting into both of these areas, I think that would be an important addition, especially when measures are being focused in that arena.

CO-CHAIR BASKIN: Well, I am going to give my specialty colleagues -- I'm a primary care doctor -- more credit than perhaps they're giving themselves. I honestly

think that the GU folks do contribute to the review of the GI measures, and the GI folks do contribute.

Because, yes, I mean, I've never -- well, actually, I have done a colonoscopy, but I've never done a cystoscopy. But there's -- many aspects of what we discuss aren't actually -- the knowledge of the actual procedure itself isn't really so important. We have our colleagues who are the specialists to be able to tell us that.

But to be able to review evidence that an assessment improves health outcomes, I don't think that's specialty-specific, the ability to be able to review that evidence and decide whether it meets certain levels of criteria. Knowledge of whether there's other evidence available, yes, certainly that's an issue. And knowledge about any details that you think are appropriate, we have the ability to ask each other that.

Is it ideal? No. Obviously, if

we had 25 GI measures, we could have had an all GI group, and the other way around, for GU. And there's no doubt that that would have been a better way to go if it were as practical. But I do still think there is some tremendous value in two specialties essentially representing themselves and then helping with the other. I think there was more contribution than, perhaps, people give themselves credit for.

Johannes?

MEMBER KOCH: I'll second Philip's point. I think that there is a value to having other ways of thinking about it. I'm not particularly clear that GI and GU per se. I think that the GI measures, having primary care, surgery, is really valuable. I think that for hepatitis C, you have to have an ID person.

I mean, there are people that bring a diversity of thought to it. And yes, we are all academically trained. We

understand how to evaluate processes. But in terms of contributing, I think we're another set of ears. So to the extent that there's a number that you need to vote, it feels very artificial, I have to say.

MEMBER MORTON: I was going to concur, Andy. I like the diversity in the group, and I like the fact that people bring in different viewpoints. I think finding the right mix is always a tough thing, as just pointed out by Johannes. What is the right mix? But we have something called Physician Practice Evaluation Committee, where we review cases for quality, and we've actually introduced different members of the hospital there.

And it's kind of interesting, because the surgeons -- if we review certain cases, there's an amen chorus that arises, like "You know, that's going to happen." And the nice thing is, when you have other people in the group, you go "Why does it have to

happen that way?" And I do think it's important to bring in diversity, so you don't bring in an echo chamber. Figuring out what is the appropriate mix is very important. I agree, ID would have been ideal. More primary care. All those come into play. But I like the diversity.

MEMBER BUTT: Just one recommendation, since you were asking for how the information could be presented. I think a single PDF with all these tables in it as a cheat sheet would be a good thing to have. Because I know that they are scattered around. There's a separate grading PDF document, and then there are tables within the guidebook, but there's a lot of information, and if you are just looking for a quick reference, it's hard to sort of navigate yourself.

So if there is a single PDF with all these tables in it -- just the tables. We understand the concept. We just need to reference it when you're grading it -- it

would be a good thing to have as a sort of cheat sheet.

MEMBER PELLETIER-CAMERON: Speaking from the GU perspective, making my vote which had equal impact on all these GI measures -- really, I was acting as a physician, just an educated academician, on these topics. I felt that I'm not familiar with the body of literature, and that the concept developers didn't give me enough of a rating of the literature for me to be able to make an educated vote on it.

So I was voting on information that I don't know anything about, and I'm not given anything about. So I really felt blind in that way, whereas with the GU data I'm more familiar with it. So despite the lack of developer information, I could make a vote. But with lack of information on the quality, I can't guess.

MEMBER MERGUERIAN: I, too, think that diversity is important, because you get

a different perspective from people who are not in the field. But at the same time, I think standardizing and creating a standard way of actually creating those measures, analyzing them, so everything is pretty much standardized and equal, so that we actually get the same results every single time, is important.

CO-CHAIR BASKIN: Public comment about the process and the pilot itself? So if there's anyone in the room, first of all, outside of the committee, that wants to comment on the process, the pilot, and how this may or may not have worked well, please feel free to do so. No obligation.

(No response.)

CO-CHAIR BASKIN: No takers within the room. Then we would open up the line. Operator, if you could open up the line for the public comment? And this would be comment regarding the pilot itself and how this was operationalized, and whether this flowed well,

didn't flow well, and any potential comments or suggestions.

OPERATOR: Yes, sir. All lines are open.

CO-CHAIR BASKIN: Sometimes I think that the world has ended when we're in this room. It should only be so quiet when I'm at home. Well, thank you all. I think this ends this. We're just going to go to next steps and timelines so there's an expectation before we adjourn.

MS. WILBON: I just have a few wrap-up slides to make sure we're all on the same page as we depart from each other. So the next stage, I think everyone's fully aware now, will be discussing reliability, validity, which is within the scientific acceptability of measure properties criterion. And then the usability and feasibility.

We do have dates set for Stage 2, and I think maybe given some of the feedback we might see how we can arrange some of the

overlap of stuff. We'll have to talk about it. But anyway, please just save all the dates on your calendar for now, and we'll be in contact about further information on that.

We'll be taking all the notes that we have from today and creating a draft report that will go out for public comment, and it will likely -- we'll probably send something out to you to review, and it won't be very long, just to say that this adequately represents what we discussed, and then we will put that up for public comment.

I think that's it. Do you have anything else to add, Taroon?

MR. AMIN: I appreciate everybody's involvement and contributions.

CO-CHAIR BASKIN: One quick question. Just for when we return next time, is there interest in us having dinner together the one night we're here when we're here overnight, or would people prefer to just make their own arrangements? Obviously, people can

make their own arrangements anyway, but for those interested, are people interested who are traveling, to try and find a place and all have dinner together? Or would you rather split up in your own groups?

It seems like there's enough people that we can at least offer that, and just ask people ahead of time so we know about how many people. And then we could -- obviously, there are enough places around. We could find something.

Okay, I just wanted to know if that was -- another comment?

MEMBER BUTT: I was just going to make another comment about Stage 2, and I was just thinking about it right now as this was flashed up. Maybe it would be a good idea that, as we give the feedback to the measure developers -- right now it's sort of unstructured -- that perhaps we could structure it according to the feasibility, reliability, those things. Because there are

lots of observations that I would have liked to plug into those sections that they would then have as specific items. So maybe formalizing that portion of it in terms of what is coming up in Stage 2 would be a good idea, since we're reviewing these and have a lot of observations which don't fit into this stage, but it would give the measure developers very specific feedback that would prepare them better for Stage 2.

MS. WILBON: We actually will be providing them, in the sense that it will be structured in a handout, that they'll get a checklist from us and say "These were the things that the steering committee suggested." So we have been documenting what those things are. But I like your suggestion of kind of structuring it in that frame of the criteria. Thank you.

CO-CHAIR BASKIN: Well, knowing that if I keep asking for comments, you'll keep giving them, I'm not asking anymore.

(Laughter.)

CO-CHAIR BASKIN: The meeting is adjourned. You can comment amongst yourselves or with me, if you want.

(Whereupon, the meeting was adjourned at 1:50 p.m.)

H

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Gastrointestinal Endorsement

Before: NQF

Date: 08-28-12

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter