THE NATIONAL QUALITY FORUM
+ + + + +
IMAGING EFFICIENCY STEERING COMMITTEE
MEETING
+ + + + +
WEDNESDAY
FEBRUARY 24, 2010

The Imaging Efficiency Steering Committee met in Suite 600 North of the Homer Building, 601 13th Street, NW, Washington, D.C., at 9:00 a.m., Scott Gazelle and Eric Peterson, Co-Chairmen, presiding.

PRESENT:
SCOTT GAZELLE, MD, MPH, PHD, CO-CHAIR
ERIC D. PETERSON, MD, MPH, CO-CHAIR
MICHAEL BACKUS, MEMBER
JACQUELINE A. BELLO, MD, FACR, MEMBER
STEPHEN V. CANTRILL, MD, FACEP, MEMBER
CARL D'ORSI, MD, MEMBER
TROY FIESINGER, MD, FAAFP, MEMBER
HOWARD FORMAN, MD, MBA, MEMBER
RAYMOND GIBBONS, MD, MEMBER
RICHARD GRIFFEY, MD, MPH, MEMBER
LASZLO MECHTLER, MD, MEMBER
PATTI RAKSIN, MD, MEMBER
DONALD W. RUCKER, MBA, MD, MEMBER

GAVIN SETZEN, MD, FACS, FAAOA, MEMBER
REBECCA SMITH-BINDMAN, MD, MEMBER
ROGER L. SNOW, MD, MPH, MEMBER

Neal R. Gross & Co., Inc.
202-234-4433
PRESENT: (cont.)

KIRK T. SPENCER, MD, MEMBER
ARTHUR STILLMAN, MD, PHD, MEMBER
JUDY ZERZAN, MD, MPH, MEMBER

HELEN BURSTIN, NQF
HEIDI BOSSLEY, NQF
IAN CORBRIDGE, NQF
SARAH FANTA, NQF
ANN HAMMERSMITH, NQF
KAREN PACE, NQF
T-A-B-L-E  O-F  C-O-N-T-E-N-T-S

Welcome, Introductions, Brief Review of Day 1 ............ 4

Steering Committee Review: Head/Spine CT/Pulmonary Measures

Review Group 1 - NQF #IEP-005-10. . . . . 7
Vote. . . . . . . . . . . . . . . . . . . . 45
Review Group 2 - NQF #IEP-008-10. . . . 48

Review Group 3 - NQF #IEP-006-10. . . . 80
Vote. . . . . . . . . . . . . . . . . . . . 89
Review Group 1 - NQF #IEP-012-10. . . . 113
Vote. . . . . . . . . . . . . . . . . . . . 143

Steering Committee Review: Cardiac Measures

Review Group 2 - NQF #IEP-011-10. . . . 144
Lunch . . . . . . . . . . . . . . . . . . . 183
Review Group 1 - NQF #IEP-015-10. . . . 185
NQF #IEP-017-10. . . . 234

Review Group 2 - NQF #IEP-016-10. . . . 240
Review Group 3 - NQF #IEP-014-10. . . . 271
Adjourn . . . . . . . . . . . . . . . . . . . 303
9:23 a.m.

CO-CHAIR GAZELLE: We are going to
shuffle the order around a little bit just to
accommodate schedules and whatnot so we are
going to do in the following order the next
three; 5, 6, and 8. We've already done 7.

Eric and I are going to split things up today.

DR. BURSTIN: Dr. Raja from the
Drug Administration will join us at 9:30 which
is in about five minutes.

CO-CHAIR GAZELLE: Okay.

DR. BURSTIN: I'm going to have to
leave about 11:00 for about a half an hour or
an hour. I just wanted to introduce you to
Heidi Bossley who many of you know from her
prior job at PCPI and is now going to NQF as
a senior director for performance measures.
She will be helping Ian out while I'm gone.

CO-CHAIR GAZELLE: Okay. So we
all are waiting for Dr. Raja to join us. We
can probably go ahead and start.
Ian, you look like you have another announcement.

MR. CORBRIDGE: I just want to make a quick announcement. For transcription purposes if you can really just make sure you state your name when you start to speak and, I guess, talk loudly. Some individuals are having a hard time hearing certain areas of the room.

As well as try to keep down some of the side conversations because the microphones are picking up all the different conversations and I think it's very difficult for those individuals listening online as well as for transcription purposes. Those are just two things to keep in mind.

CO-CHAIR GAZELLE: And then as we start with number 5, which is Appropriate Preliminary CT Imaging of Pulmonary Embolism, a lot of the discussion about the feasibility, I think, is going to be similar to tomorrow so we can probably -- to yesterday, sorry - we
can probably recall that discussion and have an abbreviated version of it if we need to today in the interest of keeping it moving.

DR. BURSTIN: One more point of clarification that Ian mentioned to me this morning that we should have been more clear about yesterday. As we look at those criteria, importance to measure a report is now a must-pass criterion so if the measure isn't passing on importance, we don't need to actually do the rest of the materials. So keep that in mind, if it looks like a measure is not making -- I mean, most of the ones yesterday made it on importance.

CO-CHAIR GAZELLE: Making it on importance meaning it has to be all high or has to be high in the middle?

DR. BURSTIN: You basically just have to say it's not low.

CO-CHAIR GAZELLE: So the majority of people are not giving it a low.

DR. BURSTIN: Exactly.
CO-CHAIR GAZELLE: Right. Okay.

DR. BURSTIN: It worked fine yesterday but in case anything comes up like that today, you can just stop and move on.

CO-CHAIR GAZELLE: All right.

Let's start with number 5.

Gavin, are you the primary reviewer?

DR. CANTRILL: No, I am.

CO-CHAIR GAZELLE: Okay.

DR. CANTRILL: Number 5 is Appropriate Pulmonary CT Imaging for Pulmonary Embolism. This is from the group at Brigham. The conditions for consideration have been met. Under d, however, the testing will be completed within 24 months. Was there an obligation to have it done in 12 months?

DR. BURSTIN: We have just changed our policy but we'll work with whoever that is to get that done.

DR. CANTRILL: I did notice that.

CO-CHAIR GAZELLE: Before we go,
could you give us the big picture of the
description of the measure, the numerator and
denominator and then go through your
evaluation, if you could, please.

DR. CANTRILL: Sure. The issue,
quite honestly, is overuse of CTPA or what
some people call CTPD, CT angiogram to rule
out pulmonary embolism which essentially has
displaced the Q scans almost universally. The
concern is now that it is so easy to order and
the results are much less ambiguous than VQ
scans used to be so people ordered them willy
nilly.

In terms of the numerator, we are
looking at people that fulfill certain, I
think, relatively reasonable criteria for
being at risk for PE. The denominator is all
CTPDs that are done. This is for a single
patient visit so time is really not an issue.

Did I give you what you want?

CO-CHAIR GAZELLE: Just so
everyone knows, the enumerator is high
clinical probability of PE, lower or intermediate clinical probability and a positive high-sensitivity D-dimer, low probability and a positive non-high sensitivity D-dimer, or an intermediate clinical probability and no availability.

DR. CANTRILL: I was going to go into that detail actually when I covered that area but thank you for that.

CO-CHAIR GAZELLE: Because I think not everyone has read them recently. It's good to go through the specific details.

DR. CANTRILL: So in terms of the Importance in the Measure to Report, this is a relatively common entity that we treat in emergency medicine with some estimates from one in 500 to one in 1,000 patients that presented with to the ED, presenting with pulmonary embolism.

In terms of the 1(a) criteria, I gave it a C. I think they do address that.

In terms of 1(b), the Opportunity for
Improvement, this is something that is documented in the literature but also something that we talked with most people in teaching programs and something we struggle with on a daily basis in terms of inappropriate ordering of diagnostic studies. They give several citations concerning the performance gap. I gave that a C.

1(c), Outcome or Evidence to Support Measure Focus, a lot of this deals with a guideline that was actually from the European literature, although it's felt that most of it, in fact, would apply to American practice as well. There has been some discussion of that in the literature as well. This builds on several other protocols in terms of evaluation of patients that may be at risk for PE in terms of trying to segment them into being high, intermediate, or low risk. Given that I gave 1(c) a C as well.

Overall for 1, I gave it a yes
with the rationale that it is an overused study and sometimes an improperly used study so there is certainly room for improvement.

In terms of 2, Measure Specifications, 2(a), again, the numerator is the number of patients who are also in the denominator who have a documented indication consistent with guidelines prior to CT imaging. What they have done and, again, this is based mainly on the European guidelines which gathers a lot of other studies together.

As was mentioned by Scott, it's a high clinical probability of PE, a low or intermediate clinical probability of PE and a positive high-sensitive D-dimer. The third is a low clinical probability and a positive non-high sensitivity D-dimer.

As an aside, the D-dimer, which is a clinical lab test used to segment the population, they come in two flavors, the high sensitivity and the low sensitivity which should really be discarded. I don't know any
place that uses low sensitivity anymore.

   It's just a troublemaker, but they
had to deal with that so that is why you see
reference to both high and non-high
sensitivity D-dimers. Then the fourth is an
intermediate clinical probability with no
availability of high sensitivity D-dimer.

   In terms of the classification of
high or low or intermediate there are several
algorithms that can be used. There's Wells
criteria. There's modified Wells. There is
the Geneva criteria, modified Geneva,
simplified Geneva. These are relatively well
known.

   Obviously there is not a lot of
universality in terms of the agreement here to
emergency positions. There is a safety catch
here. This says, "Clinical probability can be
determined by a structure prediction tool for
implicit judgment." I draw attention to that
because I think that is very important.

   Someone says, "I don't do all that
stuff but I think that this patient has a high probability of PE, that, in fact, would be reasonable justification to do the study. I don't think that's necessarily a bad thing because it's very hard, I think, with studies and with protocols such as this to forecast 100 percent certainty. We are imperfect so this gives, I think, a reasonable outlet for that.

In terms of the target population, it's adult patients age 18 or greater. They do have exclusions for unstable patients which, again, is quite reasonable. You have a patient that comes in hypotensive with sudden onset of chest pain after an eight-hour airplane flight. You don't need to go through an algorithm. We need to get a study so those people are excluded.

The data source collection, the data will be from the medical record. They make the statement that these can be easily recorded either electronically or on paper.
Again, as was mentioned by Scott beforehand, this suffers from many of the same shortcomings that we discussed in great detail yesterday so that is certainly an issue for here as well. I gave it a P, a partial, for the measure specification, again, because of the data source limitations.

The 2(b), Reliability Testing, I gave that a C because of the number of studies that we have over the years dealt with in terms of CTPA.

The Validity Testing, I gave it a P because they have ongoing validity testing now. We have already addressed the point but not the issue.

The Exclusions, again, we talk about the shock and the hypotension as being exclusions so I gave that a C.

Risk Adjustment, not necessary. Gave that an NA.

2(f), Identification of Meaningful Differences in Performance, they really don't
address that so I gave that an N.

Comparability of Multiple Data Sources, that's an NA. Disparities is an NA. Overall, I gave it a P because the validity testing is ongoing.

Usability, I think I gave that a P and it will be appropriate for public reporting.

3(b), that's NA, Relation to other NQF-endorsed measures. There are some venous thromboembolism measures but, again, they really are more complementary than anything else.

Harmonization is NA. Distinctive or Additive Value is an NA. Again, overall I gave that a P for the entire area.

Feasibility, 4(a), is a P, how are the data elements that are needed to compute measure scores generated? Again, overall 4(b), I gave an M, minimal, because of all the issues we discussed yesterday in terms of extracting this data.
The same is true for the 4(c), the Exclusions. Susceptibility to Inaccuracies, Errors, or Unintended Consequences, I think they do address that and I gave that a C. Data Collection Strategy/Implementation I gave that a P. Overall, I gave this an M because I think, for reasons we discussed yesterday, it is a significant issue.

CO-CHAIR GAZELLE: Thank you.

Thanks very much.

Are there other comments from the review group?

DR. SETZEN: Gavin Setzen. Actually, Rich we've had little bucket conferences and discussed some of the different protocols that we were reviewing and have no additional comments.

DR. GRIFFEY: I have a couple of things I would like to add, though. Just a few issues that I think bear discussion. One of those is in trying to -- this is certainly a problematic area. I think it's an important
area. I think there's opportunity for improvement.

I think the data, as I understand it, demonstrates that implicit judgment on the part of the physician is a good one for intermediate and high probability patients when you look at the PIOPED II study. Structured review is certainly helpful, particularly in the low-probability patients.

I have a little pause with respect to the intermediate probability patients here and putting those patients in the numerator with positive D-dimer just because of the concern that I don't know that there's a preponderance of evidence, or a big body of evidence showing there is a big margin of safety for those patients.

If the prevalence of disease in that group is in the range of 13 percent or so and the negative likelihood ratio is .13, then you're kind of pushing up against the edge of where you want to be in terms of getting down
the number of patients you would miss.

I believe that you want to get
down below 2 percent because the risk of
testing starts to introduce risks of contrast
induced nephropathy and other issues.
Technically, a number of studies, and I know
the European recommendations are that you can
do this.

The science is there to
demonstrate that you can use D-dimer in the
intermediate group but I feel that for quality
measures the science should be black and white
with a lot of data supporting it.

DR. SMITH-BINDMAN: Richard, I'm
sorry. This is Rebecca Smith-Bindman. Are
you saying that sensitivity of this algorithm
is not high enough that they are going to miss
some PEs in this group of patients?

DR. GRIFFEY: Yes. That's my
concern.

DR. SMITH-BINDMAN: Do you have
any idea, I don't know this literature very
well, what kind of ballpark we might be expecting to miss PEs? What is the literature using there as the algorithm suggest?

DR. GRIFFEY: Well, I think that if you're saying 1 percent --

DR. SMITH-BINDMAN: So that the prevalence in the negative group could be as high as 1 percent of PEs so patients who don't fit this algorithm, who don't meet CT based on this algorithm, would that group of patients have 1 percent PEs?

DR. GRIFFEY: I think that in the group of patients who are risk stratified by the tools to fall into intermediate probability for PE, the prevalence of PE can be as high as 13 percent, let's say.

CO-CHAIR GAZELLE: This is Scott Gazelle. We are concerned about that group, I think, with and a negative D-dimer. What's the prevalence of PE.

DR. GRIFFEY: Yes. So then after you have a negative D-dimer that drops
significantly and the negative predictive value of -- I'm sorry, the negative likelihood ratio is very good so it will drop you down below the threshold to where --

DR. SMITH-BINDMAN: Two percent.

DR. GRIFFEY: Yes, down before 2 percent so that people feel better about using that. I guess my concern is, well, how strong.

DR. SMITH-BINDMAN: Two percent is pretty high.

DR. GRIFFEY: But I think the counter-argument, I believe, is that testing with SPECT starts to introduce its own problems so 2 percent may be a reasonable threshold.

I would just like to see that number as low as possible. While I feel good about it in the low risk group, I want to voice a little hesitation there and I would like to see very clear numbers before we make that a quality measure.
CO-CHAIR GAZELLE: Scott Gazelle

again. So what you would propose is high or intermediate does not require a positive D-dimer? Is that what you're proposing?

DR. GRIFFEY: Yes. I'm saying that low requires it.

CO-CHAIR GAZELLE: Low requires the D-dimer and high and intermediate clinical judgment alone. You wouldn't require structured Wells or Geneva or anything?

DR. GRIFFEY: Well, in this measure it recommends that you use structure.

CO-CHAIR GAZELLE: But it doesn't require it.

DR. GRIFFEY: It doesn't require it but it recommends it. I think kind of along the lines of the discussion we had yesterday with respect to the head CT I think there is going to be toggle in effect.

I think that this is going to encourage the use of a structured approach and there will probably be spillover into the
intermediate range but I just think why not
start conservatively. I think you will
achieve the same thing by starting with just
the low prob group instead of putting both of
them in there. That's just my personal
advice.

DR. CANTRILL: I would support Dr. Griffey's assessment. Also I would wonder is
there a way to potentially simplify the
numerator? It's very complex. I understand
the problem.

I mean, sometimes clinical
medicine isn't simple but, by the same token,
could there be anyway to streamline this
because I've got four very different
conditions that I have to think about
individually. This goes, from my point of
view, usability to teachability. If we could
maybe make it easier, there might be something
that we might have better compliance with.

DR. BURSTIN: I think Dr. Raja just joined us. I didn't know if you wanted
to direct any of those questions to him.

CO-CHAIR GAZELLE: Are you on the phone, Dr. Raja?

Not yet. So what would you think about instead of high is better, if we just used as the numerator the patients with a low clinical probability and a negative D-dimer and then you are trying to minimize that.

DR. GRIFFEY: And if you said negative high-sensitivity D-dimer.

DR. CANTRILL: And those would get a CTPA.

CO-CHAIR GAZELLE: Right. Yes. Then what you want is the lowest number event.

DR. BURSTIN: Could you repeat that again? So the local --

CO-CHAIR GAZELLE: It seems that the discussion has said what they are really interested in is identifying patients who have a low clinical probability and a negative D-dimer that we should not be doing.

DR. CANTRILL: I would say a
negative-high sensitivity.

DR. RAJA:  That's going to be big.

CO-CHAIR GAZELLE:  Oh, here we are.

DR. RAJA:  We're not actually sure how many patients are actually high sensitivity D-dimer. Places that we've surveyed are but if we can find out whether or not people are using the high-sensitivity D-dimer and a negative and their low probability. I think that is great to use it as an over-used measure as well if that is the intent.

CO-CHAIR GAZELLE:  Scott Gazelle again. That would simplify the measure and address the concerns that the review group as raised.

DR. GIBBONS:  Ray Gibbons. I just want to point out a potential problem with that which is if I don't measure the D-dimer and now I just have low probability and I do a CTPA, that person will not be in the measure
as overuse.

DR. RAJA: Right. You guys have already gone over the initial reason why we wanted to actually review all CT scans for pulmonary embolism and then figure out which ones met these criteria.

If you actually did look at all CT scans done for pulmonary embolism and then went through and looked at whether or not they did measure D-dimer and/or had low clinical criteria you would actually catch them.

CO-CHAIR GAZELLE: So the modification could be that the numerator is patients with a low probability and either no D-dimer, no high-sensitive D-dimer or a negative high-sensitive D-dimer.

DR. RAJA: Or a negative D-dimer. Right.

CO-CHAIR GAZELLE: Would that be acceptable from the measure developer standpoint?

DR. RAJA: I think it would give
us the same outcome which is exactly the
overuse that we want to make.

DR. BURSTIN: It would probably be
helpful if they could actually give us some
data back before they make a final decision
that this is potentially a condition and have
them respond with some data and to actually
give us a sense of how different it's going to
be.

CO-CHAIR GAZELLE: Howard and then
Holly.

DR. FORMAN: Howard Forman. Are
we able to make certain that the D-dimer is
actually ordered and viewed before the study
is made just to avoid getting into a gaming
situation? I want to make sure that we are
not setting ourselves up for --

CO-CHAIR GAZELLE: You would say
that the D-dimer has --

DR. FORMAN: The results have to
be in the order somehow.

CO-CHAIR GAZELLE: Yes. Well, at
the time the CT was performed.

DR. FORMAN: You would like it in the order, though, just because it would be hard to verify it otherwise.

DR. CANTRILL: Steve Cantrill.

Clinically I'm not worried that because it behooves them to have the results back before they go to CT. If I get the results back after CT I'm screwed. I screwed myself.

MR. BACKUS: How fast do you get labs back?

DR. CANTRILL: Well, if I'm going to order the lab for a reason I wait for that result. I don't just send off labs to send off labs.

DR. FORMAN: All right. Don.

DR. RUCKER: I like the measure. Don Rucker. Sorry. I like the measure. I guess the challenge, I think, clinically though is actually on the positive D-dimers we get D-dimers on lots and lots of people now. Unfortunately they are often positive for
reasons that are, I think, just plain ill-defined.

Then we are sort of boxed into doing the CT and that is where I think the big misuse and overuse is on these low pre-test probability D-dimers that are positive.

That's, I think, where the big spend is.

MR. BACKUS: But if you don't want to act on the study, then don't send the study.

DR. SMITH-BINDMAN: It doesn't say you have to scan.

DR. FIESINGER: In the hospitals I've worked in every time the D-dimer is positive an angiogram is ordered. I've dealt with this on rounds with residents multiple times doing research on the false positive rates and we ended up scanning a whole lot of people.

DR. GRIFFEY: Richard Griffey. I mean, that's a separate quality measure.

Don't send a D-dimer in patients who don't
need one. The other caveat, and I don't want
to complicate things, but there are other
rules like the PERC rule that will essentially
identify the patients who you already have
identified as low risk to make them very low
risk to the point where you don't even need to
send a D-dimer. Those patients wouldn't fall
into the denominator because not only would
you not get the study, you wouldn't get the D-
dimer or the study.

DR. RAJA: Absolutely. That's a
great point but I think it's a little bit
outside the purview this quality of people.

CO-CHAIR GAZELLE: Carl and then
Mike.

DR. D'ORSI: Carl D'Orsi. This is
an overuse measure so once you get that
number, what size is ideal to you, 2 percent,
4 percent? Zero is ideal but what would you
accept?

CO-CHAIR GAZELLE: I don't think
we need to specify a threshold for an overuse
measure. Is that correct?

DR. D'ORSI: Okay. So 50 percent.

It doesn't make any difference.

CO-CHAIR GAZELLE: No, it would just be public reporting.

DR. D'ORSI: Okay. Fine.

CO-CHAIR GAZELLE: You'll sort of regress to the mean and the mean will move direction.

MR. BACKUS: I had two questions. Yesterday we talked about CT of the head for non-significant trauma or non-penetrating trauma. We talked about potential improvement of 30 plus percent in an organization via the measure. Do we have any sense as to what the potential improvement would be here? I mean, if you take out essentially the highs and the moderates and all that, and I guess I would couple that with the summary of evidence of high impact where it says about 1.5 percent of the patients in the ER get a CTPA and I'm wondering if you
take that down to the ER doing 50,000 visits a year, you know, that's 750 or two a day.

If you take out the medium or the moderate and high probabilities, I don't know if that whacks out half of it. Am I essentially looking at kind of one study in the ER a day and then I just wonder if we're in the significance.

I don't know the incident. You know, based on what they say here 1.5, I think people showing up in the ER because they hit their head is way more common than low indication of PE.

CO-CHAIR GAZELLE: Perhaps we could ask the measure developer, if you had even from your own institution some information on the number of low clinical probability negative D-dimer patients that are undergoing CTPA.

DR. RAJA: That's a good question. That study is actually going on right as we speak. We don't have any preliminary data.
I'm sorry. It's been about two months but we will have data for you very soon.

DR. FORMAN: I will just -- I'm sorry. It's Howie. Just anecdotally one thing is the so-called triple rule-out study which is effectively a non D-dimered low risk patient who has chest pain and radiology now has the capacity in about 90 second to do a CT thoracic arteriogram, a CT coronary angiogram, and a CT pulmonary embolism study all at once, 90 second.

It only cost about the same price as a Jaguar. It's true. I'll tell you, if you want to do it for any reason at all, this is a good measure for that because these are all theoretically low risk for PE. They are definitely high risk for other things.

CO-CHAIR GAZELLE: Triple rule-outs in that setting would count the numerator of this measure then?

DR. FORMAN: I believe so.

DR. GRIFFEY: We would need to
specify that specifically because there are other -- sorry, Richard Griffey. There are other means you can gain this in a way if you were going to get a dissection protocol CT rather than a PECT. I mean, most places will protocol those differently so you have to pick between them but there is this triple rule-out.

DR. FORMAN: This is like a trend that is just taking off and it's frightening to me because it just seems like the floodgates can be opened in no time.

CO-CHAIR GAZELLE: Kirk.

DR. SPENCER: No, that was -- Kirk Spencer. My exact comment is how do we catch these with this measure? I assume we want to catch the CTs ordered for PE but I think some of the CTs now are being ordered for chest pain above the diaphragm, particularly between the chin and the diaphragm. I think we do want to catch those but how we prove that it was ordered for PE and not --
CO-CHAIR GAZELLE: Scott Gazelle.

I would assume that they are going to catch them by CT coding for the CTPA study so that it wouldn't matter why it was ordered. It just matters that it was done without a possible D-dimer.

MR. BACKUS: The medical record -- this is Mike Backus. The medical record usually doesn't carry the coding. Does it? Usually it gets put on at billing later if we're going to go back and do a chart extract.

DR. GRIFFEY: This is Richard Griffey. If you are getting a triple rule-out but your indication is rule out the section.

MR. BACKUS: They need to be a little bit more clear, I think, about how they will capture the event.

DR. RUCKER: Don Rucker. I think for the triple rule-out stuff, I really see that as a separate measure. It's rapidly evolving technology. I think it's just a separate deal than the sort of the PE.
Those people come with a different history fundamentally, I think, than the triple rule-out patients come. I would just have that as a separate measure. I don't think I would try to glob this onto that in any way, shape, or form.

DR. CANTRILL: Steve Cantrill. It may be a little bit early to do that but I think Howie's point was very good and I think you need to look down the pike because I can see this. It's just the CTPA that is being used but, "Boy, I have chest pain." Or, "I had chest pain three years ago." Triple rule-out.

DR. FORMAN: So what I understand, and maybe some of you know better, at some institutions this is rare. It's rare for now because the radiologists are not turning in the coronary imaging. There is a turf battle that is dividing them. As people have already told me, "Oh, you'll be fine doing the coronary angiograms." I see this equipment
exist in most institutions. Using it is the next step.

CO-CHAIR GAZELLE: So my sense is we're saying that the so-called triple rule-out should be excluded. We could ask the measure developer to exclude that explicitly from the measurement.

Are you still with us?

DR. RAJA: I am, and I completely agree. I think they should be excluded. I think there are too many other things that come into play when deciding whether or not somebody needs a coronary angiogram; family history, smoking, other factors that we just can't exclude and include in this measure so we'll definitely add on an exclusion to the triple rule-out.

I do, however, completely agree in that it's going to be a big deal and it's ramping up and there is definitely some quality measure that needs to be developed for it.
CO-CHAIR GAZELLE: Okay. Rebecca.

DR. SMITH-BINDMAN: This is Rebecca Smith-Bindman. I have a quick question just about what John mentioned. Of the distribution of the indications for the low prob PEs, how many have a positive D-dimer and is that sort of close enough to what you guys have developed this for that you could understand the risk of that group and how it compares to the intermediate versus low risk group?

DR. RAJA: I'm sorry. Let me try to understand the question correctly.

DR. SMITH-BINDMAN: Low clinical probability, low positive D-dimers versus low probability negative high sensitivity D-dimer. What is the difference in prevalence of PEs in those two groups?

DR. RAJA: So I, unfortunately, don't have the data from the studies on me right this second but there is actually enough of a difference that could be adopted by
international guidelines.

DR. SMITH-BINDMAN: So the positive is high enough that it's equivalent to the intermediate or close to that group?

DR. RAJA: Exactly. And they are both high enough that they need to be --

DR. SMITH-BINDMAN: Right.

DR. GIBBONS: Ray Gibbons. As we discuss this triple rule-out issue, I want to second the point that Howie made, there is some publicized insurance industry data from the Chicago area that shows that this is now a dominate theme in terms of testing in the BD.

I just want to express a concern that if we ask the measure developers to take that out of the numerator the potential unintended consequence here will be to increase triple rule-out ordering because that becomes the acceptable now non-measured rather than CTPA.

I say that because I've seen
patients who have presented the same way to
the ED and gotten one study on one occasion
and the triple rule-out on another and it's
the same patient with the same presentation so
we may theoretically think, and I agree with
the comment over the phone, that there are
different issues that come into play but in
the real world not necessarily the case.

CO-CHAIR GAZELLE: It seems to me
that the easier unintended consequence or
easier way out is just to say intermediate
clinical probability by inflicted clinical
judgment because we don't require a structured
evaluation so if I really want to order a
CTPA-gram on somebody I just say, "Aw, I
really think they are an intermediate clinical
probability," and then they're not counted.
My concern about this measure is not requiring
Wells or Geneva.

DR. SMITH-BINDMAN: These cost a
lot more radiation.

CO-CHAIR GAZELLE: Troy and then
DR. FIESINGER: Were the triple screens ordered for the CT pulmonary angiogram or was it a separate CT?

CO-CHAIR GAZELLE: No.

DR. SMITH-BINDMAN: No. It does discount --

DR. FORMAN: We can --

DR. FIESINGER: When you do a surface CBT code would you pick up the CT pulmonary angiogram code or would you not pick it up?

DR. FORMAN: You do.

DR. FIESINGER: Okay. I mean, you still would tease out if they are doing the CT pulmonary angiogram and if they did it for chest pain that still is overuse and it still would pick it up, the numerator, with the same search approach. It would get put in there, it may not be a bad thing.

DR. SMITH-BINDMAN: CBT code --

DR. RAJA: At that point it would
be retrospectively after the CBT code has already been put in rather than at the point of the ordering which is what we are suggesting.

DR. SMITH-BINDMAN: You can't tell them apart. Which is why the CT --

(Simultaneous speaking.)

CO-CHAIR GAZELLE: One point of clarification for the measure developer. Were you planning to identify the exam that, in fact, a CTPA-gram was done by a search of the billing records that claims the CBT codes or from a search of the medical records and the charts?

DR. RAJA: That is a very valid point. We were actually suggesting I think at some point CBT codes so I apologize. I just misspoke. You're right. We would have to go back and do this retrospectively so, you're right. In a triple contrast stand we would actually pick up the CBT code. I'm just reviewing the documentation now.
DR. SPENCER: I'm sorry, Kirk Spencer. So that's what I was trying to say. They are implicitly included unless you're going to say on medical chart review it looked like that was also ordered with these other two so we are going to take them out. Is that how you want to do it?

DR. RAJA: So if we do them with the CBT codes and review them you are absolutely right they will be included. What we would have to actually do then is actually look for -- in order to exclude them we would have to look for the other two scans to make sure they were triple rule-out scans.

DR. SPENCER: But I'm still proposing that they stay in because whether they did the other two or not, if the PE test was ordered for not a good indication for suspicion of PE --

DR. SMITH-BINDMAN: Right.

DR. SPENCER: And in our hospital we also are getting a lot of pressure from the
radiologist saying, "Look, no, these are very
different studies. You know, thick. I'm
going to protocol differently between the
three." I don't know how much true triple
stuff is going on.

They will do one and we'll go back
and kind of say, "Hey, can you take a quick
look at some other structure," but if people
really are ordering, I would propose that
because they ordered apparently a PE-gram
without a good reason for a PE-gram they be
included.

DR. FIESINGER: That's what I was
trying to say.

DR. RAJA: That's true. We could
leave it in.

CO-CHAIR GAZELLE: Leave it. So
then the other question I have for you as a
measure developer, how are you going to
determine the level of clinical probability
present at the time, especially when it can be
implicit judgment?
DR. RAJA: Right. That is going to require specific documentation. Just like with other quality measures, that is going to require specific documentation by the physician.

Either they would have to include the criteria for the Wells or the Geneva or with their implicit clinical judgment or they would simply have to say, "The Wells criteria was met. They were intermediate probability and so I obtained a d-dimer," or, "I did not obtain a d-dimer."

CO-CHAIR GAZELLE: So this gets into the discussion we had yesterday. If you have a order entry system it's easy to capture that for the measure. If you don't have an order entry system for a site to participate in this measure, you would have to do, what, manual chart reviews?

DR. RAJA: You would have to do manual chart reviews and the physicians would have to know that whenever I order a scan for
a pulmonary embolism I need to document what my clinical indications are, which is really something they should be doing anyway.

CO-CHAIR GAZELLE: All right.

Other questions or comments?

Judy.

MS. ZERZAN: Judy Zerzan. I think all of this discussion you mentioned gaps before, that this is a huge gap. I see something certainly on the horizon. CT scanning is going crazy so I just want to explicitly say we need this as payers.

DR. BURSTIN: Actually, I already wrote down that, including a research recommendation from this group to keep an eye on the measure for the triple rule-outs.

CO-CHAIR GAZELLE: All right.

Should we vote? Okay. Yes, it's with conditions. The conditions are, that we've entirely modified this, with the consent of the developer, to become an overuse measure where the denominator is unchanged. The
numerator is now the low clinical probability and either no high sensitivity d-dimer or a negative high sensitivity d-dimer. We are ready to vote.

MR. CORBRIDGE: Before we do vote we would like to see if anyone is on the line or any public comments from here in the room. Okay. Thank you.

CO-CHAIR GAZELLE: All right. So we are voting on the importance characteristic or importance score. How many people would give it a high? You got a number? How many people would like to give it a middle? One. And how many people would like to give it a low? One. Okay. That means we have 18 people in the room. We'll keep an eye on that for the next vote.

All right. For the separate category.

MR. CORBRIDGE: Sorry. I'm trying to figure out how many we actually have --

CO-CHAIR GAZELLE: Eighteen.
DR. BURSTIN: There is 19 in the room right now so somebody didn't vote.

CO-CHAIR GAZELLE: Okay. Let's vote again. Was there a second medium?


CO-CHAIR GAZELLE: For Scientific Acceptability of the Measure how many people have high? Okay. How many people have middle? Eleven. That should be no lows. Any lows? Usability?

DR. CANTRILL: The total is off.

CO-CHAIR GAZELLE: So that's 20. All right. What are the highs again?

DR. BURSTIN: It was actually ten. Just so you know, this meeting has finally convinced us we are about to order those little hand-held things.

MS. ZERZAN: We've adopted new technology now that we're here.

And for feasibility. High? No highs. Medium? Five. And low? Fourteen. Okay. Now we're voting to either recommend for endorsement or not recommend for endorsement. Who would like to vote to recommend for endorsement with the conditions as stated and time-limited? Sixteen. And who would like to vote against recommending for endorsement? Three. All right.

So we can move on now to measure Number 6 which is Appropriate Head CT Imaging in Adults with Acute Traumatic Headache. Who is the primary reviewer of this one?

Judy.

DR. RAKSIN: Actually we discussed it and ours are sort of prepared. I was going to discuss another one.

CO-CHAIR GAZELLE: Which one,
number 13? No, we were going to discuss 8
next because of Howie. We could do 8 first,
I suppose, and then do 6 and 13 if you'd like.
Why don't we do 8. So we will do 8 next which
is Appropriate Cervical Spine CT Imaging in
Trauma.

Howie, you are the primary
reviewer. Start, please, by just summarizing
the measure and stating the numerator and
denominator.

DR. FORMAN: Okay. So this is
very similar in some ways to our CT imaging of
the head in the setting of trauma. The
numerator is using the evidence-based
guideline or using the tested rule and there
are two tested rules but the one that they are
using here and the denominator is all patients
that are presenting with neck trauma. I just
want to make sure I'm actually -- I'm not
saying that right, I know.

Just to point out for the moment,
it is very similar to yesterday. The one big
distinction is a lot of the evidence which this is based on is based on radiography, originally radiography as opposed to based on CT. This has been, again, a long-standing measure and CT of the cervical spine has been available, of course, not nearly as broadly used, for over 20 to 25 years in the setting of trauma.

The other important thing to mention about this just to keep in the back of your minds as we work through this is in the sense of unintended consequences this may actually increase the use of CT imaging of the cervical spine because a lot of patients that would fit the criteria for appropriate use of CT and cervical spine are appropriate for cervical spine radiography right now.

Although I realize it has nothing to do with the purview of this group, I would just point out that from a cost standpoint, from a real cost standpoint CT cervical spine imaging is cheaper than CT radiography
primarily because only probably less than 5 percent of all patients that are going to get a cervical spine CT are not getting a head CT already so from true cost. I'm not talking a societal cost. I'm not talking Medicare cost. This is something the payers can think about at another time but the true cost of doing this is de minimis. The patient is on the table. The extra time for the scanning is about 30 seconds so the true cost of doing this is de minimis.

From a real economic perspective if the unintended consequence occurs, it's not a bad thing. The radiation risk is no different. The technical feasibility of doing it is actually easier. Cervical spine radiography is more difficult to interpret even if it is compensated less. On every economic count I would say that we shouldn't worry about that but from on the ground, what would it mean for payers it could be a problem.
DR. SPENCER: But from a payer point of view if the technologies are used together 75 percent of the time they are going to be bundled soon enough anyway.

DR. FORMAN: I would hope so.

DR. SPENCER: That is exceptionally clear.

DR. BELLO: Cycle.

DR. FORMAN: What was that? Okay.

DR. SMITH-BINDMAN: The interpretation is going to be bundled?

DR. BELLO: No. No interpretation is going to be bundled.

DR. SMITH-BINDMAN: Right. This is Rebecca Smith-Bindman. The cost in terms of paying a radiologist to read a CT is not comparable to reading a cervical spine plain.

DR. FORMAN: For no good reason but I agree.

DR. SMITH-BINDMAN: For no good reason but that is huge.

DR. FORMAN: I'll be at Medpac the
next two weeks.

CO-CHAIR GAZELLE: Let's go through the whole summary of the measure first and then we'll have a discussion, if we could.

DR. FORMAN: So anyway, getting back to that, so that is the main caveat. Everything else that we're talking about, the evidence here is very comparable to the head CT so I'm just going to run through this and hopefully do it a little more efficiently than yesterday.

So starting on Importance to Measure report, I believe it meets completely in terms of the magnitude that Dave indicated. I can tell you from experience that the use has wholly gone up in the last few years and considerably so. There is a considerable amount of evidence supporting that.

The Opportunity to Improvement, 1(b), I also believe is completely met in the sense that there are clear benefits and there is comparable data to the head CT that there
are a good number of cervical spine radiographs and CTs that are currently being done that are inappropriate and could be excluded.

Outcome or Evidence to Support Measure Focus, I think, as pointed out, there are the two large studies, the Canadian study and New Orleans or NEXUS criteria, both of which have been validated, both of which count to be considerably sensitive and specific, relatively speaking, with the Canadian rule being more sensitive than the NEXUS criteria and more specific, according to this but, again, applied in radiography.

The rating of the strength, the quality of the evidence is strong and the only real controversial issue is this issue of radiography versus computed tomography. Was the factual criterion importance to measure report met? I would say yes.

Scientific Acceptability of Measured Properties. We talked about how it's
structured and what the rules actually are.

In the summary, there is one typo where it
talks about four criteria -- five criteria and
it's four but other than that. On page 4 it
just mentions at one point the five criteria
patients require. It should read four. But
that's not critical. Again, measure
specifications, I think, meet completely under
the category of Scientific Acceptability
Measure Properties.

Then under Testing Analysis
Reliability Testing we really don't have the
CT studies. We have the radiography studies.
So that's why we're listing that as partially
met.

Exclusions justified primarily on
the basis that they hadn't been tested in
populations and primarily excluding children
under 16. I think there were pregnant women
and others in that category as well and over
65. Overall, to what extent was the criterion
scientific acceptability measured properties
met? We gave it completely.

CO-CHAIR GAZELLE: Scott here interrupting. The question of testing is not testing of the criteria, it's testing of this measure and this measure is a composite of all the different criteria.

DR. FORMAN: Okay.

CO-CHAIR GAZELLE: I would say it hasn't been tested at all, if I'm interpreting it correctly.

DR. FORMAN: Okay.

DR. BURSTIN: It is only eligible for terminating.

DR. FORMAN: Okay. Thank you.

Then, I guess, under scientific acceptability then we go back to partially met. Then, under usability I think, in the absence of having had testing and knowing the exact same problems that we discussed with other measures where we were required to either parse from the chart or use a computerized position under entry, it all...
going to come back to the same basic issue. Here I would say minimally met or partially met. I'd say minimally.

So there is an existing measure with NQF which is the cervical spine radiography measure. The measures that are not harmonized now would be harmonized after testing, I presume.

DR. BURSTIN: They are not harmonized now? They wouldn't be harmonized after testing, so could you elaborate on why they're not harmonized?

DR. FORMAN: He wrote that they are not harmonized and I accepted that on its face.

DR. BURSTIN: We can ask him that.

DR. FORMAN: Okay. So the Distinctive or Added Value I think it really does considerably update where we actually are compared to where we were in terms of imaging in the emergency room.

I do think for anyone that has
experienced it you see that even though the economics have not yet adjusted appropriately, everybody is getting cervical spine imaging if they are already getting a head CT and there is no good, proved reason to not do it other than the financial concerns because the radiation can be minimized, the lux in the radiography, the speed and the risk is that you do the head CT and you do the cervical spine imaging, it's inadequate and you have to bring the patient back again. There are real good clinical care reasons to want to do that.

Under feasibility, again, getting back to the issue of how you would be able to capture the measure, you would be able to capture the data in order to provide data at the center or individual position basis so, I think, well, at the start of the summary is partially met or minimally met depending on how you look at it. I think that is the biggest hurdle once again.

I do think, much like the cervical
spine, much like the head CT issue, the individual sheet can be filled out and I do agree that even if people are going to figure out how to game the system, so to speak, by choosing certain categories it's going to force them to have to choose a category nonetheless and presumably most people are going to actually be honest so hopefully it will actually have the effect of requiring some use. I think there is excessive use out there. Overall I think that's it.

CO-CHAIR GAZELLE: All right.

Jacqueline.

DR. BELLO: Yes. Jacqueline Bello. I was the second reviewer and I just wanted to make a few additional comments.

First, to note that on page 3 under 1(b), Opportunity for Improvement, "Studies have shown that a decrease in cervical spine imaging goes up to..." And we are left without how much it might be reduced from that. I'm sure that the --
CO-CHAIR GAZELLE: Can we pause for a second? I want to make sure the developer is aware of what we are talking about here.

DR. RAJA: I am and I'm actually pulling up that data right now.

DR. BELLO: Thank you.

DR. RAJA: I'll have that for you in a few minutes.

CO-CHAIR GAZELLE: That will need to be added, though.

DR. BELLO: Right. Definitely.

DR. RAJA: Absolutely.

DR. BELLO: It'll come out in our discussion. And, if I might continue, I had a somewhat different idea from Howie in terms of whether it would suddenly increase the number of CTs. And that is because, at least from the experience where we do a lot of trauma and, as I mentioned yesterday, a lot of head and spine imaging, very few of the patients at our institution whose c-spine is being
evaluated for trauma aren't already getting a head CT.

Now, the converse of that is not true. If you have absolute head-only trauma and you're happy, then you just get a head plain but the number that don't already get a head CT and because they are already getting one, I'm of the impression that they are not getting their head CT and then going over to Plainville for the c-spine.

I really don't think it's going to increase the number of c-spine CTs to that same possible extent. The only other point I wanted to add is that in addition to the two rules, the Canadian rule and the NEXUS rule, there is this third track that you can take having to do with the range of motion and I think that that muddies the water a little bit compared to yesterday's discussion where we had a combination of two rules just because there are too few people with the courage of my convictions who are going to jump in there
and do these range-of-motion exams, I think.

   In terms of the section that we've had that says how translatable are the criteria, I think it's a little bit more muddy and I would just maybe ask the developers to expand on that a little bit. Other than that I was in basic agreement.

   CO-CHAIR GAZELLE: Thank you.

   DR. GRIFFEY: I have a comment.

   CO-CHAIR GAZELLE: Other comments.

   Sure, Richard.

   DR. GRIFFEY: Richard Griffey. I agree with both the assessments of the measure. The ACR recommendation for the appropriance criteria is to CT these patients which, I think, differs from our take in emergency medicine and in emergency radiology. I think it's not an inconsiderable concern that this could increase CT.

   I do agree with the authors that, look, if we're applying this standard to X-ray we should at least apply some standard, this
standard, to CT. There are two or three small papers, Blackmore and Hanson and some others that have tried to apply some science to the decision step between identifying not just the very low risk who don't need X-rays and the group that is at higher risk who need primary CT instead of secondary CT, instead of primary X-ray but the science is not really fully fleshed out there yet.

Those papers, and others, identify if you are getting a head CT already, the incremental issues of getting a c-spine, it just makes sense to do that. It's cost-effective and time-effective. In other patients who meet the criteria for some imaging, it would be a little concerning if this is viewed as a rubber stamp to, if they need imaging, go on to CT, in my view. At the same time. We have to go on some standard so I think it's imperfect but it is what it is.

CO-CHAIR GAZELLE: Thank you.

DR. CANTRILL: Steve Cantrill. I
have several concerns. First of all, this is the title. The title is Appropriate Cervical Spine Imaging in Adults with Trauma. I refer you to the ordering sheet. What is the one study you can order? Cervical c-spine CT.

This implies by definition that plain films of the neck are no longer adequate so, in fact, at least in our practice we have a significant number of patients in minor deceleration MVC with no loss of conscientiousness and no head trauma who come in complaining of neck pain. Of course they are going to get their c-spine study and then they don't need a head CT.

You get in the business, "Well, they are going for a neck. Let's get a head."

"They are going for a head, let's get a neck."

But we have a significant number that do not need their head done, so, in fact, there would be an increase, at least in our practice, in terms of the number of CTs of the neck that would be done.
I am concerned that this becomes a de facto standard and, as Richard mentioned, based on really not a lot of data. We're really not there yet. And I think the cost is also an issue. The incremental cost of the technical cost, absolutely.

The electrons are cheap and the time is not an issue but it's the charge for the radiologist that, in fact, will be borne by society as currently we have our financing apparently structured, I am also concerned about the radiation exposure. I can't speak with authority. Howie, in terms of plain films of the neck versus CT of the neck, what is the difference in terms of rads?

DR. FORMAN: Oh, I don't have a hard estimate right now. I would say comparable once you actually factor in the actual number of films that you are actually taking. You are taking typically seven images, even though it's considered to be a four view, but you end up taking somewhere
between two or three laterals.

    You take one AP both of likes and
an APO from altadontoids so once we get to
that, dealing with a lot of images through a
narrow area and basically, radiation is at
least comparable. I would say in reality it's
higher.

DR. CANTRILL: Okay. One other
issue, we talk about efficiency standards and
I've got to tell you, Saturday night in the ED
and I am just up to my eyeballs in patients
and I'm trying to get necks cleared.

    Plain films I can clear. C-
spines, CTs, that's the radiologist. So I've
got all these people laying around with hard
collars on and I talk to the radiologist who
says, "I've got 27 CTs I've got to read."

    When you talk about efficiency,
part of efficiency is throughput. In fact, I
see this, not all the time but in certain
situations, in fact, it's going to decrease
our overall departmental efficiency.
DR. RAKSIN: Patti Raksin. I just want to take the counterpoint to that and I have a question for Howie because we used to do planned radiographs on all of our trauma patients and I can't tell you how many patients then had to make the trip to CT, because they couldn't see -- my life has been made infinitely easier by the use of CT with reconstructions. I can clear c-spines much more efficiently.

DR. CANTRILL: I'm not saying it's a bad study. I'm just saying --

DR. RAKSIN: So my question is what percentage of patients do you think -- I remember my own anecdotal experience but what percent of the patients getting full radiographs have inadequate visualization that would require them to get a CT individually?

DR. FORMAN: Depends on the radiologist.

DR. CANTRILL: It can be as high as 68 percent.
DR. FORMAN: For someone like me it's rarely.

CO-CHAIR GAZELLE: Jacqueline.

DR. BELLO: Jacqueline Bello.

Just on the point of interpretation, I would propose that given the reformatted images in the sagittal plain, if you can clear a patient on the basis of a c-spine radiograph you can if you're not ready to wait for the radiologist to clear on the basis of the reformatted images. I think that that is just a practical point.

DR. CANTRILL: But remember that most of us don't have diagnostic good quality monitors. The radiologists won't let us have them.

CO-CHAIR GAZELLE: Don.

DR. RUCKER: Don Rucker. I think that Patti's point and Steve's point, the other dynamic here, and I'm not sure how this interprets it, seemingly many, if not most, of the plain film c-spines we get, there is a
hedge on the report that says, "If you're really worried about this, please get a CT scan."

DR. BELLO: Based on the data.

DR. RUCKER: You know, we as ER docs are left with these little time bomb thank yous. I'm sure, Howie, you don't do it.

DR. FORMAN: I don't. I'm in the private sector, though.

DR. RUCKER: Right. Places where I worked have been major university kind of hospitals and that is the norm to get those hedge kind of things. If you're really concerned, get a real test. This measure, I think, you know, in terms of where it has to be targeted, I think is in many ways not entirely clear to me.

Is it sort of more of an ER, is it more of a radiology? If you're putting that boilerplate on your reports, or your house staff is putting that boilerplate on your reports, that's a problem.
CO-CHAIR GAZELLE: Kirk.

DR. SPENCER: Kirk Spencer. Can I back up a second? Way back. Is the measure to decrease the number of people who don't need any imaging of their spine, in which case X-rays and CTs should be together, or is it to decrease the number of people that get CTs instead of c-spines? What's the heart here?

CO-CHAIR GAZELLE: My sense is it's not a comparative CT versus plain film. It's a measure that the people who are getting CT of the c-spine are getting it appropriately.

DR. FORMAN: That's why it needs to be harmonized as well. The measure is as applicable -- this is Howie -- as applicable to CT cervical spine as it is to plain radiography. The only reason we are talking about the comparative issue, I think, is because I raised the concern about the unintended consequence which I personally don't think, from my true economic point of
view as opposed to from a financial payer point of view, is concerned.

DR. SPENCER: Again, I'm sorry. I still don't understand. Are we trying to not image the c-spine of people who don't have a clinical indication?

CO-CHAIR GAZELLE: This is only about CT.

DR. SPENCER: All right. Then why is it -- okay.

CO-CHAIR GAZELLE: Richard.

DR. GRIFFEY: My recommendation is --

CO-CHAIR GAZELLE: I'm sorry.

DR. FORMAN: If there is already a measure about c-spines, then why don't we recommend to change the measure to imaging? Then you cover any modality and achieve the effect we're trying to get.

DR. BURSTIN: And just to remind people, the prior measure was actually from Harbor View Medical Center and it was the
percentage of patients undergoing cervical
spine radiographs in trauma who do not have
neck pain, distracting pain, neurologic
deficits, reduced level of consciousness, or
intoxication.

DR. BELLO: It's really the same
thing.

DR. GRIFFEY: I mean, if there
were another measure that was -- I don't
really understand exactly what harmonization
is but if there were a measure that allowed
for -- I like NEXUS or CCR. I like making
those different options available, which also
I believe includes the range of motion so I
think that's where that comes in, if you made
those sort of the standard for any imaging or
any modality.

DR. FORMAN: Although let's be
clear. Let's not include MR in this thing.

DR. CANTRILL: If I could -- Steve
Cantrill. If I could just speak to endorse
that. My suggestion is going to be to change
the title to Appropriate CT Cervical Spine Imaging. Richard's suggestion really solves the much bigger issue. If we just call it cervical spine imaging, you can do whatever you want. If you want to do plain films, you can plain films. If you want to do CT, you can do CT but we're saying, no matter what you do, you've got to have an indication.

DR. SPENCER: Kirk Spencer. From the narrow people, are there indications to do one or the other in trauma?

DR. CANTRILL: Not in these radiologist reports.

DR. SPENCER: Not in this group of patients. Okay.

DR. GRIFFEY: If you have a hard neurologic injury, then it's going to be an indication for an MRI. I mean, is that what you're asking?

DR. FORMAN: No. We're leaving MRI out of it. We're talking about, between radiography and --
DR. GRIFFEY: Between primary CT and primary X-ray, the Harbor View group is sort of at the lead, and, like I said, the science is really not there yet.

CO-CHAIR GAZELLE: Jacqueline and then Roger.

DR. BELLO: Jacqueline Bello. So the one time that there would be an indication for one instead of the other is if it's clear that this patient is going to the OR to be instrumented and the surgeon wants the CT.

DR. GRIFFEY: I wouldn't put that in this level.

DR. BELLO: I'm not putting it in here. I was answering the question, is there in the world of neuroradiology when one is indicated over the other. So that's the issue.

DR. CANTRILL: Just a comment. I think this is an example of a situation in which our historical practice of telling the radiologist what to do is a grave mistake. We
should just be telling him what the problem is. We say, "We need imaging of the c-spine," and then let the radiologist figure out what's the best way to do it. It should be bundled and paid that way.

DR. FORMAN: That's a great payer idea.

DR. RAKSIN: Patti Raksin. Jacqueline is defending my brethren. This is one case where a neurosurgeon may, in fact, need a CT scan and an MRI and an angiogram before going to the operating room with a patient.

DR. CANTRILL: My comment incorporates that thinking. It's the same.

DR. RAKSIN: But the radiologist is not going to tell us what the appropriate study is. How do I know I don't need a myelogram?

CO-CHAIR GAZELLE: That is sort of outside of this. I had a couple concerns about the measure. One is I had to read the
numerator details about 10 times before I could understand the cascading "or" statements, particularly 2(a)3(2)(b). Maybe it's simpler to people who do ED imaging but, "None of the following risk factors that allow safe assessment of range of motion."

DR. RAKSIN: Especially the "sitting in the ER."

DR. BELLO: That goes to my point.

CO-CHAIR GAZELLE: "Patient found sitting in the emergency department." I mean, at some level, everybody's done that -- I have a concern on two levels. One is, these things are not really clearly defined. I mean, what is delayed onset of neck pain? What does it mean to be found sitting in the emergency department? Moreover, I think it's going to be very difficult to determine if they exist or don't exist, all of them, from a medical record review. I think that is a real problem with this measure, personally.
PARTICIPANT: It would require a separate sheet just like the other measures that we have discussed.

DR. BELLO: No. But then you need another one for the range of motion.

CO-CHAIR GAZELLE: And definitions for all of these sort of vague things. I'm concerned about that aspect of the feasibility of the measure based on that.

DR. RAJA: If I may interrupt just a second on that. You bring up a good point. You are talking about two different things. Number one, the feasibility and collection of data which I agree is going to require a separate form.

Number two, as far as the standard division of these criteria, they are all part of the Canadian cervical spine rules which have been used now for about eight years or nine years.

Just in October of 2009 the Canadians published a article in DMJ in which
they actually used these rules in 12 centers in Britain and actually found that they were not just feasible but they actually reduced imaging overuse.

This is by a wide variety of emergency conditions, PAs and trainees, and found that even though the criteria are somewhat vague, they are actually able to be used by a wide variety of people -- including, actually, now, they've recently done a study with paramedics, actually, that have found that they can actually use the rule as well. I agree that they seem vague but they have been used, and it's been proved that they can be used.

CO-CHAIR GAZELLE: No, no. That's not what my point was. My point is not about the NEXUS or the Canadian c-spine rules. My point is about your 2(a)3(2)(b).

DR. RAJA: That is actually part of the case.

CO-CHAIR GAZELLE: I understand
that and so, if you would let me finish, what
I'm trying to say is since it's already part
of the Canadian c-spine rules, why does it
need to be in this measure separately
specified?

DR. FORMAN: You need to specify
Canadian c-spine rules without all --

CO-CHAIR GAZELLE: They already
are in the two fours above.

DR. RAJA: You're right. They
don't need to. If we can simply say,
"Canadian Cervical Spine Rule," then that's
fine.

DR. FORMAN: That is the second of
the cascading "or" statements and then it's
followed by a restatement of part of the
Canadian c-spine rule. If you just take that
last one out, it clarifies the measure.

DR. RAJA: That's true.

DR. CANTRILL: Steve Cantrill. I
have a question. What would be involved and
since we already have the previous NQF
proposal in terms of limiting for plain films
and now we have one for CT, how do we go about
merging these together? Mechanistically what
do we have to do here? I think it's really
simple and it's an extension of, really, the
one that was passed in 2008.

DR. BURSTIN: Yes, I agree. I
think the simplest approach would be if we
could talk to the folks offline at the
Brigham. I think the simplest approach would
be for us to approach the Harbor View folks
and have them talk to these folks and see if
they can come up with a measure that actually
reflects cervical spine imaging broadly.

I think that is the simplest
approach. We could table that discussion
until a follow-up conference call so that this
could get sorted out.

CO-CHAIR GAZELLE: Should we then
not vote on this?

DR. BURSTIN: It's up to you. I
think it's fine to defer that until you have
CO-CHAIR GAZELLE: Could we take
sort of a straw poll to see if everyone is in
favor of that approach?

DR. BURSTIN: Sure.

CO-CHAIR GAZELLE: Is everyone in
favor of that approach, that is, taking it

Now we are moving onto -- that was
just number 8. We decided to go to 6 and then
13 is the order we're going to have. I don't
know what we're doing in terms of scheduled --
do we need a brief break or should we push
through and let people who need a break for
food or other things go?

All right. So we're going on to
number 6 which is Appropriate Head CT Imaging
in Adults with Acute Atraumatic Headache.
That is from Review Group 3. Who is going to
take the lead on describing the measure?

DR. RAKSIN: The three of us
discussed both of these measures but -- Patti
Raksin -- I will present measure number 6 which is entitled Appropriate Head CT Imaging in Adults with Acute Atraumatic Headache.

The numerator in this case is the number of denominator patients who have an American College of Emergency Physicians indications or head CT. The denominator is the number of patients with acute headaches who are undergoing CT.

In terms of 1(a), do we believe that this met completely the criteria? We believe this is an important area for research. However, they said that the primary aim was to specify a corporate criteria and that is not what this measure is actually set up to do. This will simply identify individuals who meet existing criteria outlined in a single guideline.

In terms of 1(b), opportunity for Improvement, as the group pointed out, using the Goldstein study 98 percent of patients were determined to have a benign cause to
their headache. Fourteen percent of those patients were in that study and about five percent eventually had what they called a pathological diagnosis. They also pointed out the utilizations varied widely even within their one medical center from 5.8 percent to 11.5 percent.

We did, however, ask the developers this question. They called this acute headache but they define that as less than 14 days. Unless there is a neurologist here we take exception to that. This is really talking about subacute if they are going to stretch it out to 14 days of symptoms. For 1(b) we gave them a partial for those reasons.

Things start to get a little bit hairier when we get to 1(c), the Outcome or Evidence to Support the Measure Focus. We haven't actually talked so much about this, but when we are developing these measures for consideration, I think with the strength of
the evidence that is being used as a basis for that measure, it's quite important and this is a case where the evidence base is not very strong. Their entire measure is based on this American College of Emergency Physicians Guideline which is really a consensus statement.

As they themselves say, they are using Class 2 evidence to make a Level C recommendation and they have their own internal system so I'll just tell you, Level C in their system is, "Consensus recommendation based on incomplete, conflicting, or preliminary evidence."

Right. This is my point. I have a little bit of a problem in using that evidence base.

CO-CHAIR GAZELLE: We should pause and have a discussion of the importance because if it fails there, we don't need to go on.

DR. BURSTIN: Because the evidence
focus is required to pass number one.

    DR. CANTRILL: Actually, we have

many Level B recommendations here. They are

not all Level C.

    DR. RAKSIN: Right.

    DR. CANTRILL: In terms of

patients presenting to ED with headache and

abnormal findings in a neurologic examination.

And then patients presenting with new sudden-

onset severe headache and then HIV-positive

patients. Those are all Level B

recommendations.

    DR. RAKSIN: In the body of this

proposed measure, and if you go back to the

document, I agree with you, part of it is

listed as a Level B recommendation but then

they have made an internal mistake in the body

of this document, because they have cited --

    DR. CANTRILL: I have the

document.

    DR. RAKSIN: I have the document

as well.
DR. CANTRILL: I've got it right here. I've read you from the Level B recommendation.

DR. RAKSIN: I understand that but it's the entire way in which they've gone about generating the recommendations that were made in that document that is the problem here. It really is nothing more than a consensus statement.

CO-CHAIR GAZELLE: What I would propose then is we should hear from the other members of the review committee.

DR. BURSTIN: We've talked about it and we agree.

CO-CHAIR GAZELLE: Then we could have a brief discussion and vote on importance. Then if we agree that it's important, we'll continue on, but this will simplify it. Are there other thoughts?

PARTICIPANT: So the three primary reviewers, it's your view that it doesn't meet essentially the clinical standard?
DR. CANTRILL: Again, I disagree.

Level B recommendations are based on strength of evidence of Class 2 studies that directly address the issue. That is reading from the definition of Level B recommendations. What I read you before were Level B recommendations so I am confused about how we are saying these are consensus recommendations.

CO-CHAIR GAZELLE: Judy.

MS. ZERZAN: This is Judy Zerzan. So they say they are Level C and they include everything. If they had perhaps picked out the things that there was better evidence for, then you could say it, but as this stands covering the whole spectrum of headache I think it's inadequate.

Also, they state at the end in their rationale for using this guideline is that it's the most recent literature review published after August 2006. In my mind that is old and I don't know if there are other studies that support that but I thought this
is primarily based on expert opinion and that
is a very low level of evidence.

DR. CANTRILL: Do the authors have
any comment about that?

DR. RAJA: You both brought up
very good points, and if it says 2006, I
apologize, that must have been a typo. The
clinical policy came in in October of 2008.
You are both right in that a lot of this is
Level B evidence but there is some data behind
it. We have Class 2 studies that have
actually recommended a majority of these
points.

You are right in that we should
have excluded some of the Level C steps but
rather than make an individual decision on our
part, we deferred to the guideline of this
national body that underwent a whole lot more
review than the two of us could actually do so
we kept their guideline as a whole. If you
think that it might be more acceptable if it
simply came back as just the Level B
recommendations, we are happy to do that as well.

CO-CHAIR GAZELLE: Helen, that would be more than a revision because that would require a discussion of the whole committee.

DR. BURSTIN: It's just not clear from reading it at this point what's in and what's out. I don't think you have enough information to know how significant a change that is in terms of which elements of the guideline are B versus C, I guess is the question.

I don't know if you have a good sense of it, Steve.

DR. CANTRILL: Well, we would have to go through and actually line them up.

MS. ZERZAN: The next one that we deal about this that is prepared by CMS, I think there is a very nice job of outlining all of the different guidelines about imaging that was much more thorough and complete when
CO-CHAIR GAZELLE: Then why don't we --

DR. MECHTLER: There are multiple guidelines out there with a headache consortium, the American Academy of Neurology. I would like to have seen the ASAB on this article. They did reach out to specific neurologists without naming them and headache specialists but there was no consensus. There was no support. I'm not sure what recommendations were included in their guidelines but I would like to see this because this is a gray area although it's an emergency room issue, as you know, headache and neurology.

I would like to see a consensus from multiple groups, and change the guidelines accordingly. Because those guidelines that were recommended are a tad different than the guidelines that the American Headache Society has supported.
CO-CHAIR GAZELLE: So unless there are any more comments in favor of this, should we just vote to recommend or not -- I mean, on the importance issue? Okay.

On the importance criterion how many people in the room would like to give it a high? Okay, zero. How many people would like to give it a middle? Looks like three. How many people would like to give it a low? Fifteen. And one abstention.

So we can move onto the next measure which is measure 13, same topic. Measure 13. Who is going to be the discussant?

MS. ZERZAN: This is Judy Zerzan and we all discussed this. The summary of this measure is, it's developed by CMS using their claims data, like one of the measures that we looked at yesterday. We actually think when you look at the numerator and denominator, that what this measure is, is about inappropriate CT scans.
It's not clear from their brief description of the measure that that's what it is but in the numerator is people that got a CT scan that had reasons and they actually give excellent detailed reasons about diagnoses of those who wouldn't need a CT.

Going through our ratings, in 1(a) the imaging of people with headache is absolutely a big problem and growing so we gave that a C. Moving on to 1(b) in the Opportunity of Improvement, they give a range of the data that they have used that has a ratio ranging from zero to .8 so quite a wide range. It seems like there is a lot of variation and there is a lot of variation by state. It seems that there is opportunity for improvement so we gave this a C.

Moving on to 1(c), the Relationship to Outcomes, we gave this a partial because one thing that we were concerned about is, we agreed that there is inappropriate use and over-imaging of
headaches but if anyone needed imaging it's people over 65. That is often an indication and so while there is probably some overuse in this population, it is much smaller than in the general population so we weren't clear. They don't measure the outcomes of people that wouldn't be scanned by this and if there would be unintended consequences. I think it's about in that one study about five percent of people have pathologic problems so our concern with this is, could it have unintended consequences so we gave it a P. They did nicely review all the guidelines and the evidence and sort of summarized that, so overall for 1 we gave it yes, the threshold criteria was met. They nicely laid out the evidence some of which is not super strong and some of it is based on experts but on the whole all of the evidence is sort of in the same direction so they put it together.

Moving onto number 2, the Measure
Specifications, we gave this a C. In the numerator is the number of ED visits with the diagnoses noted in the denominator that had a CT scan and they just sort of flipped through. They have an extensive list of ICD-9 codes that do that. It is not stratified or risk adjusted and a better quality is a lower score, so fewer CTs in the not-indicated population.

Moving onto Testing and Analysis for 2(b). We gave that a C. They developed this measure with 100 percent Medicare fee-for-service sample for 2007 and then tested it on the 5 percent sample.

The validity testing, 2(c), we gave a C. 2(d), the Exclusions Justified, we also gave a C. They got their own technical expert panel that reviewed this twice and we thought that that was reasonable.

Risk adjustment is N/A.

Meaningful Differences and Importance we gave a C because they did show quite a range in
this measure. They also talked about the case count needed to get precision and gave consideration to small numbers.

2g. The Comparability, was N/A.

2h. Disparities and Care was N/A. Overall for our scientific acceptability, we wavered -- some of us gave a C and some of us gave a partial just because we were worried about, are you missing things in this age group. But overall it passed.

Moving onto 3, the usability, this criteria is not in use but they have tested it. We gave it a partial because we weren't sure about it. Well, actually, we gave it a C. Never mind. We thought that one of the things that really, this would improve the usability on is that this is an area that needs to stimulate more PY and this measure used in a younger population. If we had our druthers and could rewrite the measure we would have substituted this measure into the younger age group because we think that's
important although it's much harder to measure
those in different health systems and much
easier in Medicaid.

Harmonization is N/A. There is no
other measure similar. There are no competing
measures so that is also N/A. Overall we
thought that the usability criteria was met
and that this would help sort of push things
into looking at appropriate measures in the
younger population where overuse was a much
bigger problem than in this population.

In terms of feasibility we gave it
Cs for a., b., and c. because this is
electronic data, claims data that Medicare
has. For 4(d) we gave it a partial because we
were worried about the unintended consequences
of missing disease and we weren't sure of the
magnitude of that problem or if it was just an
uncomfortableness on our part.

For data collection it's C, it's
Claim Data. Overall feasibility is a C. I
think that's all. Oh, Recommendation. We
said yes, with added on that we want other
payers to use this, that the younger
population has the bigger impact.

CO-CHAIR GAZELLE: Okay. Thank
you. Are there any other comments from the
rest of the review group?

DR. MECHTLER: Well, it's an
interesting look at headaches, primary
headaches, because really you are excluding
all secondary headaches with neurological
deficits so you are really looking at primary
headaches. The question is, what percent of
primary headaches actually occur after the age
of 60. It's low.

Having said that, I think this
could be expanded to a younger age group that
would be academic and more intriguing as a
headache specialist. That is one issue. The
other issue is that they mention data should
be looked at requiring CT with contrast
because the indication for contrast in CT for
uncomplicated headaches with no history of
cancer, no history of infection, is relatively low. We don't use contrast with teens so that is something that needs to be looked at. Otherwise, I think I agree with Judy that this is a study that could be looked at and has some merit.

CO-CHAIR GAZELLE: Thank you.

Patti, any comment?

DR. RAKSIN: No. I have pretty much the same. The age group is the issue. If we could merge this with the other study looking at a different group, then I think we'd be happy.

CO-CHAIR GAZELLE: Okay. Thank you.

Steve, do you have any comments?

DR. CANTRILL: Steve Cantrill. Yes, I have several problems with this proposal. First of all, it attempts to -- obviously, you all are able to read this -- so, what it tends to do is cut down overuse but gives absolutely no guidance in terms of
how to do it. How am I going to cut down the number of head CTs I do? Just not CT on Thursday or maybe not CT anyone over 60? That would make everybody happy.

That's what is really lacking here. The previous measure that we talked about did have its limitations but its attempt was to give guidance to the provider of care so they can actually reasonably try to limit and decrease the number of inappropriate CTs. This does nothing of the same. This gives me a number that I don't know what the heck to do with it.

CO-CHAIR GAZELLE: Unless I'm wrong it says that it reports the number of CTs done in a series of conditions where it's implied that CT is not appropriate.

DR. CANTRILL: No, it says, Of the ED visits identified in the denominator, visits with a coincident Brain CT study.

CO-CHAIR GAZELLE: Yes, so the denominator by ICD-9 code lists all the
different commissions.

DR. RAKSIN: This is Patti. I can clarify this. It's actually a little bit tricky when you read it the first time because the denominator has a number of exclusions which functionally serve to limit things to the number of patients who really don't have a real indication for getting a CT. In that sense I agree with you that it's not as clear as the last one when you check off the box yes, there's an indication, do it.

I look at this as a first step in a QI process. This is an attempt to identify the magnitude of a problem. I think once you have identified the magnitude of the problem, then you can take appropriate steps.

DR. CANTRILL: Can we make this a little clearer and easier for people to use?

DR. RAKSIN: I don't disagree with you. I had to read it three or four times.

DR. GRIFFEY: Richard Griffey. So I guess I'm a little confused up front at how
a measure that is addressing the same issue could be deemed important for one measure and not important for another. That is the first thing. I think that these measures get at the same issue from different directions.

One is to say these are the people who would be appropriate to have a CT. The other it says get a CT in everyone but these people. Not everyone but don't get them in these people. It's kind of getting at the same thing from two different directions.

This has a feel of a utilization review is what it is and that is what sort of makes it kind if impalpable. It's hard to know how exhaustive or complete the list of exclusions is. Not just in terms of the things that should be there that aren't there but should all things there be there.

One of the evidence citations in this is the very clinical policy that was found to be lacking in support of the other measure. To me I don't feel the real
justification for one measure over the other in the way it is right now.

DR. RAKSIN: This is Patti. If I could clarify again. So the difference in our analysis is that with the first measure in order to pass that first hurdle you have to show two things. One, that it's an important clinical problem.

I think we all agree that it is.

The second part of that is whether the evidence basis to back up the measure is there and that is where we as a group had a problem with that first measure. Yes, the same clinical policy is cited but it's cited among many other documents which form the evidence basis for this measure.

DR. GRIFFEY: Which were also cited in that policy.

DR. RAKSIN: Yes.

MS. ZERZAN: Because there are points -- this is Judy Zerzan again. There are parts of the American College of Emergency
Physicians that did have better evidence than what they showed so I think this measure takes the best of all those.

It also has the U.S. Headache Consortium, the Singapore Ministry of Health, the American College of Radiology. I think while it doesn't tell people explicitly, it's not a prospective measure. It's a retrospective measure that is similar to the one that we passed yesterday. Their rates are zero to 80 percent which is a huge range.

I think probably the right answer is not this measure should be zero but it also probably shouldn't be 80 percent. Knowing the spread of that will then allow people to look at why is there variation, do more studies, and figure out what is the right rate.

CO-CHAIR GAZELLE: Carl and then Steve.

DR. D'ORSI: I'm sorry. Just a point -- Carl D'Orsi. Just a point of clarification. I'm still a little confused
and I'm very sorry. It's just not my area.

On the denominator exclusions are they sort of
the numerator over everything? In other
words, when this ratio was done with a number
greater than 1 -- be developed? I'm unclear.

DR. RAKSIN: No. This is Patti.
The denominator exclusions here are basically
most of the things that you would think of
that would give you positive findings.

DR. D'ORSI: So then why aren't
they the numerator and all the CTs a
denominator? Why was it written like this?
That's what I'm getting at.

MS. ZERZAN: We had sort of said
that we didn't really like the name of this in
the description because what this is the
number of people that should not get a head CT
but did and it's confusing.

DR. MECHTLER: The American
Headache Society and the AN have come up with
the -- if you have a primary headache with a
normal neurological examination there is no
indication for imaging. That is what has been published over the last 10 or 15 years.

Having said that, over the age of 50 is one of those red flags that could be a red flag. Once you get over 50 then it's metastases, temporary otitis and other causes of headaches that are quite significant.

If you do have a primary headache, and most of this is in a denominator that is included are primary headaches. They can be cluster migraines, episodic tension headaches, some rare forms of headaches such as exertion and so on.

They are looking at patients who have headaches over a specific age that should not get imaging and if they do get imaging, maybe take that information and find out what percent of these patients could turn out to have primary headaches at imaging.

CO-CHAIR GAZELLE: Richard. And I'm just going to remind everyone we have a tight schedule so let's try as best we can.
I don't want to cut off discussion but let's try and make the comments short and new so that we can get on to the voting.

MR. BACKUS: I think one of the potential positive things about the measure is that, again, with the Medicare population the feasibility has improved. That said, how often are these exclusion criteria actually coded?

Someone comes in and you work them all up and at the end of the day you don't find anything concerning and headache gets written on the chart but you wouldn't necessarily include other elements that would work for or against you. It sorts of relies heavily on the coding piece of it. It good for feasibility but it's also a threat to the validity.

CO-CHAIR GAZELLE: Okay. Mike.

MR. BACKUS: This is Mike Backus. I think, Steve, to your point it just tells you that the measure is high but it doesn't
tell you what to do. That is why I really contract this to the other measure because in the other measure there is, to me, as I heard the two groups debate about what is clinically appropriate and I don’t pass any judgment there but in this case what it says is if you are an outlier on one end or the other, then you as an institution go back and look at those clinical guidelines and now you as an institution figure out where your standard of practice is.

Are you happy with B level evidence? Are you happy with C level evidence? Where do you want to go. It at least points you in the right direction. To me the feasibility of the measure just becomes overriding because so many of the other things that we talked about is, you know, we've got to get a paper form.

We've got to get the physician to dole something out. We have to hope to go and get a hospital with an ED and a physician
group with an ED to come and do paper forms
for stuff that if their percent of charges
reimbursement is going to cut their revenue.
And here, essentially, we are open to
something where we can at least go get a look
at the data set for free.

CO-CHAIR GAZELLE: So you are
speaking in favor of it.

MR. BACKUS: Strongly.

CO-CHAIR GAZELLE: Okay. Other
comments. Start by saying whether you are
speaking in favor or against.

DR. SPENCER: Well, I'm clarified.
One of my problems with the CMS measures that
I'm trying to figure out -- that maybe CMS can
answer -- we have ABNs now for anything we
order through Medicare so if I order a test,
our hospital screens it.

If it doesn't pass they stick a
form in my face and say, check another box.
We can't order an adequate test without
passing an ABN so if ABNs are going to be
everywhere, shouldn't all these eventually be zero? Or are ABNs not everywhere, just outpatient, beneficiary, notary?

DR. DEHN: That is where you don't think it's going to be covered by Medicare. This is not a payment. In other words, the measure is not a payment, doesn't affect payment. In other words, the ABN --

DR. SPENCER: No. They are telling me to have the patient sign it again because it looks like Medicare is not going to cover it.

DR. DEHN: Okay. Susan may want to comment but using this measure is not related to the payment for Medicare services.

DR. ARDAY: Not at all.

PARTICIPANT: Maybe I can help. The ABNs were introduced for physicians who chose to practice medicine or perform procedures in a similar gray area. In an attempt to guarantee payment before that goes on, they would like you to certify that's
done.

On the other hand, there is no question right now for payment unless you were to do an audit for medical necessity. Your hospital is very aggressive in terms of asking you to fill out that information. To get Medicare to pay for a head scan is not an issue.

CO-CHAIR GAZELLE: Ray.

DR. GIBBONS: Ray Gibbons. This is just a point of clarification. It's about Section 2(f)3 that describes the observed data because I don't have the supplemental file. This section gives the outliers. I would like to know the median and interquartile range that was observed. I'm sure you have it. I just would like to know it because I think it's relevant to the precision.

CO-CHAIR GAZELLE: All right. We can ask them to get that while we have any additional comments.

Rebecca.
DR. SMITH-BINDMAN: This is Rebecca Smith-Bindman. I had one question and I'm not sure if it's different than Richard's or not but just the validity.

CO-CHAIR GAZELLE: Try and keep that down in the back.

DR. SMITH-BINDMAN: My question is just the validity in two ways about this measure. The codes are very specific for headache and I don't know the reliability of doctors completing those codes. My guess is not very high. My question is two-fold. I think you are raising concerns about the usefulness of this measure. I just also want to raise concern about the sensitivity of this measure for the exclusions. Are all indicated CTs going to be captured and excluded from these measures?

PARTICIPANT: And maybe if headache has such a low need for imaging we'll pay you. I just want to make sure that patients who are generally in need of a CT are
not getting it based on compliance.

CO-CHAIR GAZELLE: Thank you.

Okay. Are there any other comments from the steering committee? If not, we have a chance for comments from the measure developer if there are any. Any comments from the measure developer?

All right. Any public comments?

Is there any public comments?

PARTICIPANT: Sure can't even get a job. My own fucking --

CO-CHAIR GAZELLE: Hello? Is somebody still on the phone there? Are you still intending to be on the phone? Is there someone on the phone who would like to be on the phone still? Alright. The F word was enough.

Okay. It's time to vote.

MR. CORBRIDGE: We need to dial back in before we move forward. We just disconnected the phone line.

PARTICIPANT: Is someone on the
phone?

MR. CORBRIDGE: We can leave it open for individuals who want to --

CO-CHAIR GAZELLE: As far as I know no one is on the line. Could we start the voting? He already asked for public comment and there weren't any.

MR. CORBRIDGE: Can we dial back in quickly so we can just kind of set things up for the voting? We'll just dial in quickly.

CO-CHAIR GAZELLE: Okay. We are going to be voting on this measure we've been discussing.

DR. MECHTLER: Just a logistic. In the denominator they said primary cough heading -- primary cough headaches.

CO-CHAIR GAZELLE: Okay.

DR. RUCKER: I was wondering about that.

CO-CHAIR GAZELLE: Okay. So we are going to vote now. We are voting on the
importance criteria. Are you ready? How many people would like to give it a high?

Importance, 13. How many people would like to give it a middle or medium?

Keep your hands up until we're ready. Six. No lows then. Okay. For Scientific Acceptability how many people would like to give it a high? Two. How many people would like to give it a middle? How many people would like to give it a low? Four. Okay.

For Usability how many people would like to give it a high? One. How many people would like to give it a middle?


Now we are voting to recommend for endorsement or not to recommend for endorsement. This would not be time limited. Correct? How many people would like to vote
for the endorsement of the measure as written?
Fifteen. How many people would not vote for
endorsement of the measure as written? Four.
Okay.

DR. BURSTIN: Can I clarify? This
is with no conditions, right?

CO-CHAIR GAZELLE: No conditions.

DR. BURSTIN: No age changes or
nothing like that?

CO-CHAIR GAZELLE: No conditions.

Okay.

Now we are ready to move on to the
last of this group, the CT Pulmonary Measure
number 12, Simultaneous Use of Drain CT and
Sinus CT. The primary reviewer on this one is
from Maurice Oblan.

DR. SETZEN: Yes. Gavin Setzen.

CO-CHAIR GAZELLE: Okay. Please
summarize the measure and then take us through
the review.

DR. SETZEN: Just to give you some
background where this is coming from, given
some recent data and certainly in the last
decade with the data from CMS where there has
been five percent per annum increases in CT
imaging utilization as well as any JN data
about over-utilization, radiation risk and
cancer rates, as well as other potential
consequences from imaging of the use contrast
and false positives and things like that has
spurred a lot of the debate about
appropriateness, utilization, safety, and
efficiency.

I think that is largely what
drives this measure. I think it's a very
reasonable consideration just to give you the
numerator. The numerator is looking at
patients who present to the ED and most of
this data is 2007 claims data in hospital
outpatient or ED settings.

Patients who receive both head CT
and a sinus CT and the denominator is head CT
alone. The goal is that the lower the number,
the better the outcome. The idea is to reduce
the clinician's potential for inappropriately ordering a sinus CT scan for somebody who is being evaluated for headache and maybe getting a head CT.

Part of the rationale behind that is that your head CT which will demonstrate much of the sinus and nasal pathology as almost a screening mechanism. The same can't be said for patients having a sinus CT and evaluating the brain.

So in terms of the overall review there is a handout as well that presents some of the data similar to the headache data that was just passed around so that's useful to see. Going through the recommendation in terms of importance to measure certainly high impact and there is certainly a lot of supporting data and literature.

I think it's also an important opportunity to change clinician behavior with respect to ordering appropriate studies, lessening the potential radiation exposure and
things like that. From that perspective I gave 1(a) a C rating.

There are many good citations.
There is good evidence from the American College of Radiology, American Academy of Otolaryngology, headache panels and other consensus data that is out there in the literature in support of what constitutes an appropriate head CT and appropriate sinus CT. We'll get into exclusion criteria and so on a little later on.

In terms of opportunity for benefits as a mechanism for ordering appropriate studies on appropriate patients with perspective safety and efficiency and overall cost, I gave that a C as well.

Moving onto Outcome or Evidence to Support the Focus, 1(c). It's important to note that there are specific exclusion criteria in the study and those are clearly worked out in the literature be it the American College of Radiology, Otolaryngology,
Head and Neck Surgery and others.

For example, abnormal neuro examinations, headache worsened by valsalva, headaches awakening one from sleep, new headaches in an older patient, progressive worsening from the sinus nasal standpoint, sinus or nasal polyps, unilateral disease, concern about malignancy, neoplasm, orbital cellulitis and factors such as that are exclusion criteria which did not alter the data or skew the data or limit the potential for having concurrent studies.

Good citations in terms of evidence with respect to overall threshold in terms of importance to measure and report, I gave that a yes.

Moving onto number 2, Scientific Acceptability, the measure specifications, again, a low score being the goal with reducing the number of concomitant head and sinus CTs over the number of brain CTs alone.

2(a) would be a C. There are
certainly very specific CPT codes and those exclusions which will allow for that process, the exclusions we spoke about and, again, the denominator exclusion.

These are claims with primary and secondary diagnosis codes related to trauma, concern about potential neoplasm, orbital cellulitis, intracranial abscess. Those are very clear in the document and that becomes very helpful and important as well.

With respect to Stratification Risk Adjustment, not applicable in this scenario. The data source is claims data. It's administrative and not necessarily presenting as a burden to the ordering physician or clinicians involved.

With respect to Testing and Analysis there are proportions taken and the developers are cognizant of potential issues as it relates to appropriate coding for the ordering entity to potentially input an incorrect code or as it relates to code
modifiers so there is care taken to specifically identify and address those as well as thresholds at both ends of the spectrum in terms of meeting the minimum number of studies required to make sure this is a reliable standard that facilitates better validity.

The measure basically uses, as I said, the Medicare outpatient SAFs. The data for a lot of this is in the 2007 data summarized in the supplemental handout. So with respect to 2(c). I gave that a C as well. The exclusions are certainly justified and very specific for both the head and sinus components with respect to imaging. The 2(d) would be a C.

With respect Risk Adjustment and so on, 2(e) is not applicable.

2(f), Meaningful Differences in Performance, the summary and supporting data demonstrates a few different things including geographic variations in terms of utilization.
Certainly metropolitan areas is a much greater density in terms of utilization. I think that provides an important opportunity for improvement.

Certainly there is an interesting performance gap for those that haven't seen the supporting data with respect to inappropriate additional ordering of a sinus CT concomitant with a head CT. I think that's really the most important product where there is an opportunity for improvement. For measuring the scores and testing and current use and so on, 2(f), I gave that a C.

For Comparability of Multiple Data Sources, not applicable, 2(g).

2(h), not applicable.

Scientific Acceptability of the Measure Properties, I gave that a C for number 2.

Moving onto Usability, 3(a) in terms of Meaningful and Useful Information, again, this is claims data review. Promotes
high quality, efficiency at the end of the day. I gave 3(a) a C.

Harmonization not applicable.

3(c) not applicable.

With respect to overall usability criteria being met, I gave that a C, number 3.

Number 4, moving onto feasibility, again the data is extracted using coding by another individual other than the person obtaining the information. It's claims data facility-level information. 4(a) I gave a C.

Electronic Sources, 4(b), a C.

Exclusions, 4(c), a C.

With respect to Susceptibility to Inaccuracies and Errors, or Unintended Consequences, just the potential for miscoding including entry at the point of the ordering physician but certainly the possibility of modifiers and that is something that can be monitored and tracked and excluded if necessary.

So with respect to Data Collection
and Strategy, 4(d) was a C, 4(e) a C as well.

Also there are no incremental costs or administrative concerns other than what would be absorbed on the measurer's end. So with respect to 4, Criteria for Feasibility, I gave that a C and overall recommendation yes to proceed with the measure.

CO-CHAIR GAZELLE: Okay. Thank you very much.

DR. SETZEN: You're welcome.

CO-CHAIR GAZELLE: Other comments from the review group? Steve.

DR. CANTRILL: Just one that I think has been brought up before. Unfortunately this really does limit the population we're looking at which really limits unfortunately the impact of this measure.

CO-CHAIR GAZELLE: Because of the Medicare only issue?
DR. CANTRILL: Yes.

DR. GRIFFEY: One or two small comments. It was a little hard for me to get a sense exactly of the magnitude of the problem here. I would be surprised to learn that it's a big huge utilization problem.

DR. SMITH-BINDMAN: It looks like it's 5 percent in the numbers they quote is the median so 5 percent of CTs haven't been sinus CTs. It seems not that bad.

DR. GRIFFEY: I can't really comment too much on its use in routine practice and it's hard for me to imagine many scenarios where this plays out outside of that. I think the one thing that I think always comes into play with this kind of measure where it relies on these exclusions is, again, the documentation of them.

If someone had -- I'm trying to imagine a scenario. A concern for an abscess that then gave you meningitis so you wanted to order a head and a neck CT. If you didn't
find an abscess you wouldn't document an
abscess and then you wouldn't get the
exclusion and then you get dinged.

    I think that's a very kind of rare
scenario. I'm really not that worried about
it but that is the kind of problem you get
when you rely on coding of negative findings
to enter you into the exclusion population.

    That was the same thing with that
last measure. Unless you took the time to
document, oh, yes, by the way, they had some
dizziness, and you get that thrown out of the
measure, then you're at risk for getting
dinged for that.

    CO-CHAIR GAZELLE: Yes, Roger.

    DR. SNOW: Two words to confirm
that. Folks just don't document negative
things and they don't document in detail if
they think they've got enough to move. They
think in terms of two or three things. All
the other stuff is there and we just don't
know it.
CO-CHAIR GAZELLE: Ray.

DR. GIBBONS: Ray Gibbons. I want to express some concern about the validity numbers. It applies to any of the measures where the range is actually very small. The interquartile range is a series of small numbers so in this case the interquartile range is from .022 to .047.

I would point out that the number of cases being used only allows 90 percent competence elements of plus or minus .05. For a small hospital that means one year there is zero and the next year they are graded in the 75 percentile of the country and that's on the basis of chance alone.

I would strongly suggest that this measure needs to be reconfigured to use a much higher number of cases as a cutoff to be fair to smaller hospitals because it's statistically not valid in that range. It's just a fundamental limitation which I think doesn't negate the potential use in larger
hospitals but will cut down the number of hospitals that are in this sample.

DR. SETZEN: Right. That's a good point. When you look at the weight of the averages the standard deviation is .0020 and so where they talk about 45 cases is the minimum and then adjusting accordingly for that small facility, that will present a problem.

DR. GIBBONS: So I would suggest the number has to be enough to make the 90 percent competence limits to be smaller than the standard deviation.

MS. ARDAY: When we are doing this and moving the competence down we were taking into the account the ratio levels so it's not like we set a single minimum case count. Actually because of the nature of that for distribution and the data they are varying the case count requirements relative to what the data --

DR. GIBBONS: Ray Gibbons again.
I didn't really read this one through but I read the CABG one through and that is not what it says. It actually says that low case counts. At the low numbers they used 45. That gives you the same .05. That's not what the CABG 1 says. If this one is different, then somebody better --

CO-CHAIR GAZELLE: While we are looking, other comments?

Carl.

DR. D'ORSI: Just very quickly. This is to discourage bundling or routine ordering, for example, of the sinus because you can't clear the brain with an ordered sinus CT. Is that correct?

PARTICIPANT: A lot of the conventional thinking and data out there depending who you read, up to 90 percent of sinus headaches are actually migraine or a typical headache variant and not sinus in origin at all.

DR. SETZEN: So how often if
somebody correctly orders a sinus CT do they then go ahead and simply order a brain CT?

PARTICIPANT: Rare.

PARTICIPANT: I'm not sure. My experience has been usually if there is a sinus CT the only time the head is CT'ed there is some unusual abnormality that they pick up on the sinus CT that is intracranial. I would say that is probably an appropriate extension of a head CT.

The other way around is that really a head CT should give you the sinuses and I concur with my ENT colleague that over 80 percent of chronic daily headaches are actually migraines or chronic migraines. I think that here the acute sinusitis, I would love to see some history because are these chronic pain syndromes or acute because that information I would like to gather.

I'm not sure there is an opportunity to gather any clinical information. From what I see here you are
just looking at the acquisition of images, head and sinus, without any history, without any symptomatology in my mind this would be an important study just getting information. I'm not sure if that is feasible.

CO-CHAIR GAZELLE: Yes, Jacqueline.

DR. BELLO: Jacqueline Bello. I agree that this is a significant problem but more so in the much younger age group where the opportunity to scan through the lens is just not resisted often enough in children who have headache.

That said, for whatever group you're applying it to given their methodology, I don't understand why hydrocephalus is not in the denominator exclusions.

DR. SETZEN: I think more specifically any intracranial abnormality or neurological issue clinically presenting would be an exclusion. It's not specified as a separate exclusion criteria.
DR. BELLO: My question is why isn't it in terms of you're writing a measure. Granted it would pertain much more to the age group that I think needs this more but there is plenty of shunted adults out there whether it's for MPH or obstructed.

CO-CHAIR GAZELLE: So you're proposing a modification.

DR. BELLO: If it were to fly I believe that given their intent hydrocephalus should be a denominator exclusion if it were to fly.

DR. SETZEN: That's reasonable especially given the high percentage of asymptomatic patients.

CO-CHAIR GAZELLE: Any objections to that on the steering committee? Is that acceptable to the measure developers?

DR. DEHN: If you want to clarify that. You know, if it's acceptable we'll take care of that. This came from the experpal at locus who suggested exclusion so their numbers

Neal R. Gross & Co., Inc.  
202-234-4433
were based on that.

CO-CHAIR GAZELLE: Thank you.

Anymore comments from the steering committee before we ask for additional comments from the measure developers?

Yes, Rebecca.

DR. SMITH-BINDMAN: Rebecca Smith-Bindman. Can the measure developers just address the issues of sample size and measurability of this relatively uncommon 5 percent issue and sample size for facility levels and ability to actually come up with useful rather than just noisy numbers.

CO-CHAIR GAZELLE: So let's ask the measure developers for any comments but please address those issues if you could and then we can have final comments before we vote.

DR. DEHN: Maybe while doing that if I could refer you to 2(f)2 in the document and then response.

DR. SMITH-BINDMAN: I have it
open. These are random and variable and the sample size for this is --

CO-CHAIR GAZELLE: Okay. So let's

--

MS. ARDAY: We end up with 3,330 facilities in which our statisticians, you know, thought we had sufficient case panels to measure this. I want to remind everybody that the denominator on this is all brain CTs.

There are a lot of them so we were doing the statisticians. One of the reasons the number is low is because the denominator is all the brain CTs and then what you're looking for is the simultaneous. These are large denominators.

MR. CORBRIDGE: Well, all brain CTs minus all the exclusions.

MS. ARDAY: Minus all the exclusions. Correct.

DR. SETZEN: In 2007 -- Gavin Setzen -- there were 120,000 brain CTs done looking at claims data for patients presenting
to the ED with a headache diagnosis just to
put that in perspective.

MS. ARDAY: This is actually
beyond the emergency department.

DR. SETZEN: Head and sinus in
2007 there were 2.1 million in the
denominator, the numerator. Of those patients
who had combined head and sinus CTs, 80,000.

MS. ARDAY: Once we apply the case
counts the number and the denominator in the
aggregate is 1,909,644. The numerator, and
this is what gives you the small number.
Understand that the denominator is a large
number but 70,271.

We are working with a denominator
that is very large. The technical expert
panel because this was a debate as to whether
you narrow it and have the denominator be
defined with a primary diagnosis of headache.
Our expert panel thought we should do it with
all brain CTs.

DR. SETZEN: That factor was .037
when you do the math.

DR. SMITH-BINDMAN: I'm sorry. This is Rebecca Smith-Bindman. The denominator summing across the 3,000 facilities doesn't help me. What I want to know is the mean number of relevant patients at each facility. My guess is that's closer to a few hundred in the median if you're using a minimum of 45 and we would expect the average of this to be two patients is 5 percent. One patient more than that puts you way over the top and one patient less. We're talking about a difference of a single patient so we understand the competence interval. We want to know the reliability of your measure so we want the distribution of the number of relevant cases per facility.

MS. ARDAY: What I can share with you is they had for the facilities, and we have this in the handout because we feel we have the percentile distribution of the numerator and the denominator, at the very low
end, you know, in the denominator we obviously have 45 cases. We go up to a maximum of 4,000 cases.

At the median this is where we are rank ordering the hospitals. At the median we have 462 in the denominator. In terms of the numerator the minimum obviously ends up being zero. The maximum was 184. At median, half the hospitals, the ranked order was 15.

DR. SMITH-BINDMAN: Say that one more time?

MS. ARDAY: Fifteen.

DR. SMITH-BINDMAN: Is the median?

MS. ARDAY: Is the median.

DR. SMITH-BINDMAN: You are looking at half that are less. The half that are more you kind of get -- with the half that are lower than that.

CO-CHAIR GAZELLE: The noise in the ratio for the institutions that have numbers less than 15 in the numerator is going to be really high.
DR. SMITH-BINDMAN: So the question is sort of how many facilities would be consistently characterized from year to year assuming the exact same performance allowing for one patient to change.

CO-CHAIR GAZELLE: Okay. Other comments from the measure developers before we have final comments from the steering committee, public comment and then vote?

Other comments from the developers? Any other comments from the steering committee?

MR. BACKUS: This is Mike Backus. So 15 is the numerator. What's the median denominator, 400?

MR. BACKUS: 460, so it's 15 out of 460.

MS. ARDAY: Because the denominator is such a large number your ability to -- a group of a few cases is not going to make I think a huge difference.

DR. SMITH-BINDMAN: But isn't that what you want to do, though? You want to
judge the quality at the facility level or am I misunderstanding you?

MS. ARDAY: Yes, you do.

DR. SMITH-BINDMAN: So then --

MS. ARDAY: In other words, I think how frequently a facility does a simultaneous study, I think you'd have to have it be fairly common to have it move from where it sits in one year.

DR. SMITH-BINDMAN: That's my question. I'm wondering if a third of the facilities couldn't switch categories based on single patients. That's my question. I can't do it on the back of a napkin.

MR. BACKUS: Is your denominator adjusted for exclusions?

MS. ARDAY: After exclusions 3.7 percent.

MR. BACKUS: And then one more patient moves you to 4 so you would have 16 out of 400.

CO-CHAIR GAZELLE: It moves you
from 3.3 to 3.0 percent.

DR. RUCKER: But that is median.

The real issue is what you are doing at the 25th percentile and below. The hospital is at 45. That's what they're using. I will reiterate my objection. It is unfair to those hospitals. It's fine that the measure gives you a value of .4 which the last one did. It is not fine and it's giving you a value down at this level because the .05 decision is inadequate.

DR. SNOW: Is this another opportunity where you wait until you accrue enough cases before you --

DR. SMITH-BINDMAN: Use the measure.

DR. RUCKER: You can do it either way but the easiest thing is to just raise that number so it is valid for -- it cast aspersions on the whole measure unnecessarily by getting too much noise at the low end.

Those numbers are not going to contribute
significantly to the national problem.

DR. SETZEN: The intention is good
but the unintended consequences for those
small institutions can be lasting and very
significant.

CO-CHAIR GAZELLE: Can we --

MS. ARDAY: This is Susan Arday,
CMS. Since this is pay for reporting and not
pay for performance are you recommending that
we establish a cutoff where it's 75 percentile
and above would be what would be publicly
reported?

CO-CHAIR GAZELLE: It may be pay
for reporting for CMS but for NQF it's
neither.

MS. ARDAY: I'm hearing gained and
hearing other things where it makes it sound
like there is some cumulative --

MR. DEHN: You have to understand
that, first of all, there is no -- I mean,
okay, let's start with the base of it. This
is something that the guidelines say should
not be done so basically the numbers should be
in medicine, I know it's not perfect, zero.
They say you shouldn't overexpose people and
you shouldn't do this.

I don't see CT scanners and this
is where we are telling hospitals, they are
not being dinged. All you're saying it is
inappropriate and should not be done. We
understand there are some variability. A
hospital would have to go just to move from 16
to 20.

Well, it's four cases but a
hospital is 16 cases and the next year four
more move to the other side means that we're
assuming that only the new ones are done
incorrectly. If they now have more of those,
that means they have a bigger pool so we can't
do it just that the new one is going to change
the variation.

Furthermore, moving from 16 to 20
means that you are increasing by 30 something
percent your inappropriate use of medicine in

Neal R. Gross & Co., Inc.
202-234-4433
this case. I mean, I know they are small numbers and we're talking a two million denominator and that affects the whole picture.

DR. SMITH-BINDMAN: I don't think it's a question of importance if the validity of the measure that you're saying has been used. The example you gave was a very nice example of a situation where we have sufficient sample size so if you can just address that so it's 45 cases and one had inappropriate and the next year two had inappropriate can you really say something about the quality of that institution, or might the lack of validity of the exclusion criteria being coded in the CMS records be more important that the movement of that one case?

CO-CHAIR GAZELLE: Or even one person making a mistake.

DR. SMITH-BINDMAN: Or one person making a mistake would you want to label that
hospital as over imaging?

MR. DEHN: I think one percent or
one case really in 45 people would say --
first of all, we're not saying 2.7 is good and
3.8 is bad. That's not something we're saying
first of all. I think it wouldn't shift in
that case. I would agree that --

DR. SMITH-BINDMAN: So if it's two
or three cases out of 45, it goes to 7 percent
versus --

CO-CHAIR GAZELLE: I think we have
had discussion on this topic so let's move to
-- unless there's new points that need to be
made let's move to opportunity for public
comment. Is there anyone on the phone? Any
other public comments?

DR. SMITH-BINDMAN: No.

CO-CHAIR GAZELLE: So can we move
to voting on this now? All right.

PARTICIPANT: This is not time
oriented?

CO-CHAIR GAZELLE: No, this would
not be time limited. There were no
modifications proposed so far.

PARTICIPANT: Hydrocephalus.

PARTICIPANT: There was not a
specific sample size modification.

CO-CHAIR GAZELLE: Okay. So in the
importance criteria who would like to vote
high? Okay, none. Who would like to vote
middle? Who would like to vote low?

Helen, do we need to continue?

All right. We are done with the
first two groups then, Head CT and Pulmonary
and Mammo. We need to move onto Cardiac.

Lunch is scheduled for 12:30 but could I say
one schedule question? All the documents that
came out when we were urged to make our travel
reservations cited an end time of 3:00 and the
current agenda says an end time of 4:00. A
lot of us have flights that won't allow us to
stay until 4:00.

DR. BURSTIN: Why don't we ask how
many folks need to leave by 3:00? Okay,
that's it, so 3:00 is the end. Again, if we can't finish work we can always do it at our last conference call. It's not ideal but -

    CO-CHAIR GAZELLE: We only have six and we've got three hours so we can do it.

    DR. BURSTIN: Is that our average?

    CO-CHAIR GAZELLE: We've been going through these in about 25 minutes this morning. Do you want to do them in order? Does anyone have an objection? Anybody got to leave early? Let's do Group 2, Ray Gibbons, Measure 11.

    DR. GIBBONS: This is Measure 11, the Use of Stress echoes, Myocardial Profusion Imaging abbreviated MPI. That is SPECT nuclear imaging for those who wonder. And Cardiac Stress MRI Post CABG.

    The background for this is that there is certainly an enormous problem in terms of the rate of growth for cardiac imaging. Some of it is nicely referred to in the summary of evidence that's listed but it
is certainly well documented in the literature that in the late 1990s and early 2000s the rate of increase in cardiac imaging far exceeded the rate of increase in other cardiac conditions be it myocardial infarction or cardiac treatments such as stenting.

The latest data that was published in the American Journal of Radiology, if I recall, last year showed that that trend continues up until 2006 in the out-patient area. The compounded rate of increase exceeds 15 percent per year. There is no question that cardiac imaging has grown dramatically.

In response to that the ACC, American College of Cardiology Foundation working with various other partners, has tried to develop appropriateness criteria that attempt to indicate when imaging should and should not be done.

By way of full disclosure, I was on the very first technical panel for the very first appropriateness criteria which were
indeed for SPECT myocardial perfusion imaging.

Since that time, also by way of full disclosure, my laboratory and I have been involved in evaluating appropriateness criteria both at the Mayo Clinic and in general, and have published a paper showing the spectrum of the problem, and also elucidating some of the problems with applying appropriateness criteria, which will be evident as we discuss these measures.

So this one is an attempt to look at stress imaging after coronary artery bypass grafting, particularly in the first five years when one of the appropriateness criteria sets, stress echo, said it was inappropriate. When the stress SPECT criteria said it was of uncertain appropriateness.

The measure is basically to look at the total number of patients undergoing coronary artery bypass grafting in the denominator over the last five years. The numerator is then to look at those patients
who have undergone SPECT imaging.

   The numerator excludes certain
categories of patients and that is part of the
difficulty because the appropriateness
criteria talks about asymptomatic patients
following coronary artery bypass graphing.

Using administrative planned data that is hard
to come by so the numerator excludes a series
of patients who have certain ICD-9 codes.

   It also excludes patients who have
undergone testing within six months of
coronary artery bypass graphing. It also
excludes patients who, following SPECT
imaging, have gone onto coronary angiography
or interventional procedures.

   So in terms of the importance, I
think, I think it's an important area but I
would point out that it is not clear from the
available data just how high prevalence this
issue is in terms of absolute number. The
data submitted with the application shows
calculations across the country.
The denominator average across 3,000 hospitals is 240 and an average rate that they show of .016, that would actually come to 12,000 procedures annually compared to the total volume or procedures reported in the same document for 2007 of 789,000. That's 1.5 percent defect. That is small compared to the annual rate of growth in outpatient SPECT procedures.

MR. BACKUS: This is Mike. Is this looking at strictly hospital or is this looking at imaging if you're a cardiology practice and you have your own camera?

DR. GIBBONS: Strictly facility and strictly Medicare, too. Correct?

MR. BACKUS: Outpatient hospital.

So with respect to the Section 1(b), which reports clear variation, one of the issues in that variation is, again, the precision of the measurement given the very low rate that we are finding, a .016 across the country. The issue of timing after coronary artery bypass
grafting as patients were tested within six months or excluded and I don't understand that exclusion.

Also the rate of cardiac catheterization after those SPECT studies is an exclusion. I would question whether some of the variation is actually due to that exclusion and I personally don't understand that one either. I don't think any of those three areas, the precision, the timing for the path are opportunities for improvement.

With respect to Section 1(c). I actually would wonder since subsequent tasks are then excluding these patients whether the consequence of this would be to inspire more coronary angiograms after the unnecessary stress procedures. There is potential harm, I think, from that standpoint.

Do my other colleagues want to add anything before I move on to the next section?

Summary statement on 1 would be –

CO-CHAIR GAZELLE: Let's have your
scoring on the importance

DR. GIBBONS: Partial.

CO-CHAIR GAZELLE: Partial.

DR. STILLMAN: I have one other thing I can add in the support material.

There was a graph of the utilization rates geographically across the country. The highest -- the black ones tend to be ones in more rural states. The inference, of course, is that if you are one of those states you are doing a poorer job but it may be in a more rural environment where you don't have as many cardiovascular specialists that this might be the best kind of cure you can get.

I had a little bit of a problem with that. Even the variability, if you look at it, seems to be sort of lopsided in those rural areas.

MR. BACKUS: This is Mike Backus. In disclosure, we run a pretty significant MPI management business at AIM. I would strictly conjecture -- I would guess that because we
are only looking at the facility side and not the practice side when you get to those more rural areas perhaps the cardiology practices there are on the border of whether or not they have enough volume to own the cameras so that may go out to the facility.

My only question with the measure is, based on what we see, in the commercial population for preauthorization there is substantially much more of this interview done in the office than done in the facility. I would worry about an unintended consequence of essentially, you know, as a referring physician you have the ability to move where that exam goes.

DR. STILLMAN: Arthur Stillman again. You're dealing in rural areas where cardiac cath is available.

MR. BACKUS: So then it doesn't qualify for -- right. I understand the exclusion if you ever take six months, you might have something going on with the
transplant or the CABG so we'll take that out.
I understand why those are acceptable. It's
the same thing.

You have a cath done down the road
and you are effectively saying, "The fact that
I had a cath down the road meant that taking
that image was okay. Now I'm right at the
limits of my critical knowledge."

The stuff that I read or we talked
a lot about internally says that cath isn't
necessarily the only way to manage somebody
where stenosis has been found so I don't know
that the cath exclusion is the gold standard
there, the downstream cath.

My only concern is if we are only
looking at the facility side, you know, you
have the ability to move where those exams go
and that may be causing a data problem.

DR. DEHN: If this would help, we
certainly stipulate to the comments that were
made. One of the basic reasons that we wanted
to do this and talk about it is there seems to
be an increase number of acquisitions and practices by hospitals.

    As Mike said, currently about 80 percent, almost 85 percent of all myocardial profusion imaging is done in physician's offices, so this isn't a whole lot. We recognize that.

    What we would like to do is gather -- you may or may not approve of that for that particular reason but it would be a sham for you to think that this is a great scientific endeavor that we want.

    What we want is -- hopefully you will endorse the use of this as a baseline in which to measure what we consider to be a significant change in the way healthcare is being practiced and that much of this will be moving back to the hospital. How much we don't know but we would like to monitor it.

    That's the best I've got.

CO-CHAIR PETERSON: One question.

I'm sorry. I still don't get why are you not
calculating outpatient? Did I miss it?

DR. DEHN: It's a tricky thing, folks. I appreciate that. The mandate that this group has that we're advising is to look only at inpatient hospitals.

DR. BURSTIN: Outpatient.

DR. DEHN: Excuse me. Outpatient hospitals. We do have all the other data. We could give that to you under the table but that is not part of our mandate so we aren't allowed to give it to you.

MS. ARDAY: Initially we were given TRICIA money. TRICIA money could be spent on hospital outpatient issues but not on the broader --

DR. BURSTIN: But if you had the broader set of imaging facilities you could try whatever you want.

MS. ARDAY: That opportunity. I mean in terms of the burden of caring it would be a lot more but it would give us much more --
DR. DEHN: I think candidly if that would be the recommendation of this committee, it would certainly give us some opportunity to go back to Baltimore and say, "This is the real way to do it." Suggestions and modification from you would be welcome.

CO-CHAIR GAZELLE: I guess what we're hearing is that we have one modification proposed. Is there general support for that around the table? You may not support the measure but at least support the conditional change.

DR. D'ORSI: Is this basically to open up as much of the net as possible?

DR. SPENCER: If we don't, then we can stop discussing the measure.

CO-CHAIR PETERSON: Okay. So now we are moving one.

DR. STILLMAN: As far as the measure specification the numerator is spelled out with a variety of exclusion codes. This gets to the heart of the difficulty of
defining asymptomatic in an administrative
database. The ICD-9 codes are very broad.
They include cardiac dysrhythmias, syncope,
palpitations, orthopnea, for example, other
chest pains, abnormal ECG.

It's a very broad set of diagnoses
that are then excluded from the numerator
which, from my standpoint, I would say as a
cardiologist I cannot justify testing in
somebody with palpitations but that is when
impact is --

CO-CHAIR PETERSON: This excludes,
right?

DR. STILLMAN: It excludes them
from the measure so it's okay to do that.
It's okay to do that.

For Section 2(a) I would,
therefore, give them a partial. For testing
and analysis we have already talked about this
issue but I would point out that the median
value -- the weighted average is .016. The
median value is actually zero. No, the 25th
is zero. The median value is .008 so we are using a measure in small hospitals which is plus or minus .05 trying to measure things at a .008 level. It is not going to be valid in a smaller number of hospitals. From a validity standpoint I would give it minimal.

The exclusions I've gone through. I really do not -- I don't understand excluding people within six months of CABG. In other words, you can test them at seven months but you can't test them at five months. I know of no science to support that and it's not reflected in any national guideline that I'm aware of.

CO-CHAIR PETERSON: I suspect that having worked with these kind of curves before and actually done a study, I think the early exclusions were put in there in part for return to work and other stress testing that may be done as part of job requirements.

Six months is a broad window to get to that but I think on each of these to
compensate for the fact that they don't have asymptomatic is compensated by kicking out codes which gets you out of it. To compensate for the early testing that might be required for reasons of work they have given a window of time. I'm not justifying.

DR. STILLMAN: Likewise if you do a test that is really -- taking that person to TAP then excludes them.

CO-CHAIR PETERSON: And the compensation there is if you have a positive study or something that created a need so far as the position to actually put a patient in an invasive procedure and it in part would maybe --

DR. STILLMAN: Encourage more angiography after equivocal studies.

DR. SPENCER: This is Kirk Spencer. I agree that's just disturbing. As you said earlier, if you've got an unindicated test, that's what we're trying to prevent is unindicated CABG because they've got an
asymptomatic patient with some defect score of six. Now people feel compelled to CABG when they didn't need the stress to begin with so that is disturbing.

MR. BACKUS: I agree it's potential unintended consequence. I just think to think that a physician who, you know, a patient goes through CABG, the physician comes back and has a reason to look. Let's assume all that is excluded. It's an asymptomatic patient and then six months down the road or nine months down the road on an asymptomatic patient they come in and they do a stress test.

I don't think the incentives are strong enough for them to turn around and say, "Well, to meet an NQF measure I'm not going to take a patient who has had a CABG into the cath lab." I think --

DR. SPENCER: You're missing my point. People will cath for --

MR. BACKUS: Why would they cath?
DR. SPENCER: The range of who to cath. I mean, they're not pregnancy tests. They are not yea or nay. They are normal or outrageously abnormal and the problem is where you cut off and where you need to cath is poorly defined.

One of the reasons not to test asymptomatic patients because you don't know what the hell to do with the result once you get it. If you're in the middle, you feel a bias to cath. I'm not suggesting they are going to do it for the measure. People feel uncomfortable if they have a moderately abnormal stress test sitting in a patient's chart.

MR. BACKUS: So if your point is that then they wouldn't go to the stress test, I agree with you. The measure here, what we are essentially saying is we have excluded all kinds of things so that anybody doing these tests is potentially fairly far out in the inefficient curve. Right? So I don't know
that they are now going to say, "I was inefficient on my stress test and now I'm going to be --

DR. STILLMAN: Just from a measurement standpoint let me sort of summarize for this section by pointing out that I find the exclusions hard to justify. That includes palpitations, abnormal electrocardiogram, cardiac catheter PCI after the procedure and SPECT for stress echo performed within six months.

DR. SPENCER: 427.61 which is a PAC so I would point out that as these exclusions increase the rate will decrease but quality will not necessarily be any better and, in fact, may be worse.

MS. ZERZAN: So could we perhaps cut out some of those diagnoses that you think are less gray areas?

CO-CHAIR PETERSON: There will be at some point a fundamental decision here that we'll have to say are we comfortable with ICD-
9 codes trying to define an asymptomatic really at-risk population or not.

I do want to inject a little bit of data both for and against so you can sort of see. We have a paper that is actually accepted from our group that is going to be looking at 28,000 patients' data being linked with United Healthcare looking at this pattern of testing so it actually compliments in part what they present here for 65 plus. This is under 65.

We looked at 28,000 people undergoing revascularization through the UHC's national database 2004 through 2007. Of that this will reflect both the PCI and the CABG but I can give just the CABG number. 7,000 of those underwent CABG procedures.

Rates of testing we excluded the first three months window, I think, in ours but rates of testing from three months then onward to 24 months out after the procedure that fall within the guideline of
inappropriate.

Fifty-one percent of the patients undergoing CABG underwent a stress test so it isn't small unfortunately in this country. You would be surprised to know remarkably recorrelated those episodes of chest pain.

They happened to happen at six months and 12 months at convenient office visits to the cardiologist. I'm sure it happened to work out quite nice. The diagnosis codes that were most common 75 percent of them were 414. as the most common cause. Chest pain did account for 23 percent.

DR. SMITH-BINDMAN: What was the other one, 414?

CO-CHAIR PETERSON: 414, ischemic heart disease. You have disease. I'll keep going. There are a couple of other factors that relate to some of the things you were going through. The degree to which actually in this study places that were the highest there was 40 percent variation in use of it
across major cities. Here, Phoenix, Orlando, Dallas, Houston, Cleveland were the culprits. Shocking.

The final thing is the rate of people who undergo revascularization after the procedure itself and of those tested 11 percent underwent angiograms and of those five percent underwent repeat revascularization so 95 percent of these didn't yield any further stuff.

DR. D'ORSI: Eric.

CO-CHAIR PETERSON: Yes.

DR. D'ORSI: Carl D'Orsi. Was the numerator dissimilar to what was placed in this numerator in your study? What was the numerator?

CO-CHAIR PETERSON: The numerator were people who would have fallen in this category. It was considerably higher because this is around 50 percent would have fallen in --

DR. D'ORSI: What was the --
CO-CHAIR PETERSON: The denominator was somebody who underwent a CABG. Numerator would be people who got a noninvasive stress test.

DR. D'ORSI: Just period. That's all you were looking at.

CO-CHAIR PETERSON: In that first two-year period.

DR. D'ORSI: So we don't know how much of that was not warranted and how much was.

CO-CHAIR PETERSON: Right.

MR. BACKUS: And you did facility and office?

CO-CHAIR PETERSON: This would have been across all.

MR. BACKUS: Yes, across all.

That's fine.

DR. D'ORSI: It just says these are how many people had a stress test. We don't know how many were good or how many were bad.
CO-CHAIR PETERSON: Again, we weren't doing this to propose this as a measure but it just gives you a magnitude. There is no doubt there's a problem. The question is is claims data the way to get at the question of reading it. That's where it comes out for me.

DR. GIBBONS: I want to just point out, you know, we see this map and you can see that in terms of the cities you listed Orlando is Florida, Dallas is in Texas, Phoenix is in Arizona. All those states look good on this measure even though they are bad in your clinical studies.

CO-CHAIR PETERSON: And I suspecting that is the outpatient. Sorry. Where are we with regard to --

DR. GIBBONS: I scored for the concerns that I've listed 2 as --

The exclusions, the case counts, I'm very concerned at the rate we'll go down as those exclusions increase but, in fact, the
quality may, in fact, we worse.

CO-CHAIR PETERSON: Yes.

DR. GIBBONS: May be inverse. For

3, Usability, 3(a). I thought was partial.

3(b). I guess is not applicable because as far

as I know there are no other measures. 3(c)

is not applicable.

I do think I have a bit of a

concern here in the sense as everybody looks

at the public domain where a lot of people

think more care is better care whether people

will actually recognize that being low here is

good.

As far as feasibility, the data

abstraction issue does get a little trickier

here because of the ability to reliably code,

for example, abnormal ECG which, at least,

when I asked somebody in CMS 10 years ago that

wasn't felt to be reliably prudent as a

diagnosis.

I would also point out that there

is more in weeks issue with respect to SPECT
imaging. Many of these are coded 78464 which is single image rest or stress, so with the numbers that are proposed it's conceivable and we don't know. At least I don't know.

MR. BACKUS: You get a stress code with it but you get a 9301.

DR. GIBBONS: But that's not what's shown in the proposed measure. There is no matching that I see so some of these may be resting SPECTs which are not actually the domain of the criteria.

In the interest of time, Mr. Chairman, I personally didn't recommend it because I had many concerns.

CO-CHAIR PETERSON: Great.

CO-CHAIR GAZELLE: Other discussion of the group? Other discussion by the committee?

DR. GRIFFEY: What about the proposed modifications

DR. GIBBONS: They would require, in my view, total redoing of the exclusions
and redoing of the data to see what it looks like figuring out a new number of low end. This would be extensive.

DR. D'ORSI: Carl D'Orsi. One quickie. Why was there a bundling of stress echo cardiac SPECT MPI and cardiac stress MRI? Is there any trend where these are going out and something else is coming in?

DR. GIBBONS: There are trends with regard to patterns and utilization. All of them would fall in the same bucket of appropriateness as best we can tell.

DR. FIESINGER: I hear your problems with measure but everything you're saying details what I see where I live, it the big institutes. It's a major cost excess. It's something we've got to deal with. I don't want to see the issue die even if this isn't the right total approach. This is being grossly overused.

DR. SPENCER: Yes. Kirk Spencer. I also agree that it can't be recommended as
is. I don't want the two cardiologists on the panel saying this is not a -- measure. It would be that we don't understand it's a problem.

DR. GIBBONS: Ray Gibbons. I'll second that for the public record. I am on public record in terms of what I've written and what I've said. I think this is not a well-designed measure. Personally I don't think it goes far enough and it will cause methodologic problems for all the reasons given.

DR. SMITH-BINDMAN: This is Rebecca Smith-Bindman. I know nothing about this topic but from what you're saying you have outlined very concrete things you want to see happen and they don't seem that huge to me. You're saying you want a completely different sample and they are kind of nodding over there that they can do it with the sample.

You want some of these exclusions
which weaken the measure to be eliminated and they are kind of nodding over there. And you want some sample size corrections which won't be as big a problem once you have these other facilities. If they could do those things, can you just --

DR. BURSTIN: No, the only possibility here would be to just give them a set of questions that you want to have answered and have that measure come back with different data so don't vote on it as is I guess would be the only recommendation.

CO-CHAIR PETERSON: The other thing that's relevant here to bring up, and unfortunately I don't think they proposed it, of the ACC measures there is a very similar one you'll run into in a few minutes about after PCI use of the test, why one group proposed it in one form of data and one group proposed it with another.

The ACC does use criteria that are clinical to get to the asymptomatic population
but there are then the challenges just like we had the Brigham situation that, in fact, there are some challenges in collecting that information in current practice. I don't know how to do this in terms of order. We can finish this measure and realize just in the back of your mind that alternative potentially is out there.

We'll take a few more comments and then if people around the table want to do an extensive rewrite and revote we could do that and table it today. We have option B would be we say no even with the rewrite we would not be happy because we still don't think we can get it in the asymptomatic population by claims data when there is no reason for them to do it.

DR. RUCKER: Could we consider the PCI measure because it sounds like they'd need to be harmonized and be pretty similar anyway. Could we just consider that one?

CO-CHAIR PETERSON: They use
different data so they would need to be harmonized so they one of them in the asymptomatic population defines it and attempts to get rid of the things that might be reasonable reasons why you would order a test based on claims data.

MS. ZERZAN: This is Judy Zerzan. I'm a fan of these CMS measures because it gives a set number to what happens around the country and it's something that we can compare our data to see are we way outliers or not.

It may not be a totally perfect measure but I think the improvements would help it a lot be something that would be meaningful to my constituents. I guess I would propose that we sort of decide if we want to modify this and move on and consider the other one in isolation even though they may get the same thing.

DR. FIESINGER: Basically taking your idea the measures we looked at are both pre-op evaluation. They are very similar,
there is a lot of overlap. I would like some way to look back after we look at all of these whether it's harmonization or some other format. We are all concerned about is it including the same thing. We are all going the same direction. Let's try to be on the same bus so to speak.

DR. DEHN: I share your agony in which exclusions to add and which not to. In our committee the particular question was chaired by Pam Douglas, who you probably all know. This was pretty much what she said a lot, that she recommended. Some of these you could include and you don't have to include but they tend to smooth.

I mean, you could probably do this with none. I mean, with no exceptions at all and still have some sort of meaningful variation. Let me just say that we didn't pull the exclusions out of our ears. They are there and they are there for discussion if you choose to do it or not.
As for expanding this coverage, non-hospital based facilities would be a trip to Hollywood for all of us. To the extent that we could make this work, we will certainly work with you and harmonize. I mean, to add PCI in here would not be difficult at all.

MR. BACKUS: That's a real interesting thing that you started to get to is that we work through this and we say in CMS you are already looking at post-CABG, look at post-PCI and you look at the way the ACC wants to comment post-PCI you'll get a very quick data validation or divergence of what asymptomatic really is. You'll have two different methods of defining it.

DR. BURSTIN: If I could just suggest that perhaps we table this discussion so we can complete the discussion of the ACC measure and discussion the conclusion at that time.

DR. SPENCER: Kirk Spencer. I
I think the PCI and the CABG area also have a number of different issues. I think it's worth a brief discussion for the committee to discuss whether to get this data without any exclusion and then just consider, hey, you don't want to be -- if the range is 12 to 85 percent, if you're in the 85 percent group, that's a bad surgeon whose grafts are all going down or what are you doing?

You don't want to be in the real high range. That is almost cleaner than trying to figure out the right reasons to do it. I don't know.

DR. RUCKER: I think that would be what Eric got. That's exactly what he did.

CO-CHAIR GAZELLE: He got dinged in reviews. How low on the food chain is that? We're stuck in real life.

DR. BURSTIN: That is pretty nice.

DR. RUCKER: I understand the statistical predictive model of taking things in and out of the model in terms of, gee, it
doesn't change it. I think part of the whole NQF process is sort of having face validity to these things. I think if you put in things that just simply even at the onesies and twosies of the story in USA Today or whatever journal you've got at home -- just kidding -- you know, if they don't have face validity I think it's a very corrosive type of outcome. I think it harms the NQF process and all of this things. There is sort of entire quality metric when there are certain obviously outliers because it's just corrosive to public support for this. I think you have to be very careful having things that have face validity.

DR. BURSTIN: I will just make the point that actually, again, we want to stay grounded in this which is fair evaluation criteria. Very clearly in this last round of updating saying exclusion should only be there they have to be justified. We don't want the onesies and twosies. You really want exclusions but if you didn't have them, you
1 would significantly distort the measures. I think we are trying to get away because feasibility falls off the planet when you start adding 100 onesies and twosies. Just from our perspective I think there is some valid consideration for having a set of exclusions that in a sensitivity analysis would significantly change your result as opposed to the onesies and twosies just really based on this.

CO-CHAIR PETERSON: So the issue all comes back into how people can interpret a number. On the one hand here you can say that up to 20 or 25 percent based on how data, which is probably similar to yours, people had diagnosis that could be a very legitimate reason for testing and certainly would have been not covered by the data that supported not testing in the population.

We could argue that there would be the USA Today headline NQF says you shouldn't be doing testing in a group that should be
tested by every other criteria and you are
going to condemn grandma to test fate.

On the flipside of this is to say
that, yes, the 20 percent is distributed
generally equally and we're not looking for
100 percent on this measure. We are looking
for outlier values for the guys who --
everybody.

Just from mammography logic we
were just trying to find outlier status as a
potential marker and that would be legitimate
around this table of understanding. The
question is how does it play outside of this.

DR. GRIFFEY: I want to state that
a utilization rate is not a quality measure
and so it needs to indicate that there is some
appropriateness or inappropriateness.

While it would be great, I think
everyone would love to see the data, we also
want to know what do you make of this number
that you arrive at and how you compare with
someone else if it's not case-mix adjusted or
adjusted in some meaningful way. It's just a number.

CO-CHAIR GAZELLE: Okay.

DR. BURSTIN: Think about this over lunch and return fresh.

CO-CHAIR PETERSON: Okay. It's lunchtime.

(Whereupon, the above-entitled matter went off the record at 12:35 p.m. and resumed at 12:55 p.m.)
CO-CHAIR PETERSON: Now, the discussion and decision in consensus of our leadership here has been we are going to table this now unless there are some major issues. What we are going to do is then go on and visit the CCI measure proposed by the ACC.

The rationale for that is that it provides an alternative means for which this data could be collected at sort of more on the clinical collection of data and reasons for a test as opposed to the alternative similar to what we went through in the morning, although not completely analogous to the issues that we had with Brigham and Women measures.

The challenge will come to the ACC, as you will see, on the question of usability or feasibility. If you will turn your docket over to 15.

DR. BURSTIN: Can we first see who is on the telephone?
CO-CHAIR PETERSON: Is anyone on the phone? Okay. I will begin. Cardiac stress imaging not meeting appropriate use criteria: Routine testing after percutaneous intervention. The denominator for this is number of stress SPECT MPIs and stress echo performed. The numerator is the number of stress SPECT or stress echo studies done in asymptomatic patients within two years of the most recent PCI.

In part the measure, off the bat, is that one of them is supposed to undergo the procedure and then try to look at the number who get tested. This one looks at the number who get tested and then say how many of them have gone for an inappropriate indication. Different strategies but with a pretty similar end. We'll talk about that in a bit.

Let's go through sort of the evidence briefly. There is evidence that would support an appropriate use published by Neurology and Cardiology through an intense
process. They went through a random criteria that comes out to say that this is an inappropriate use if the patient is asymptomatic.

There is also evidence given, and I could provide more. We sort of heard it. It's even higher than 60 percent in percutaneous intervention of patients undergoing testing in the first few years of the procedure. There is a fair amount of this done.

The flip side, though, is that they list a series of studies. Ray and others have published work looking at use of it in appropriate indications. Of all the testing done this is a modest to small percentage of the total testing that is done. Depending on what you use as numerators and denominators it changes again with the total procedures done. So in the importance category I gave this a C. The 1(b), demonstrate quality problems, I gave it a C.
Measure 1(c). I gave it a C.

When it gets to the 2 measures, scientific acceptability measure, I gave it a C. The rationale here having to capture and exclude patients who have symptoms.

Reliability testing, they have done some of this based on a system in a couple pilot sites where they looked at indications for a stress test. They were able to do this but it's in a limited setting today. There is by no means a universal or even broad systems or even systems that currently exist out there necessarily but you capture point of order or point of capture rationale for stress testing.

They have done some reliability testing, got a C. The validity testing gets NA and the exclusions or outcome measures, NA. 2(f), NA. 2g., multiple data sources comparable results based on what they have done so far by chart review in a small setting. Disparities is NA.
Usability for the challenges are
the information is usable and understandable.
Harmonization I had NA but now it may become
an issue depending on what we do with the
other comparable settings for -- NA on the
3(c).

Feasibility issues they have shown
that this can be done in limited settings.
Rarely available in electronic records I would
say is an M. Exclusions, NA. 4(d), P and
4(d), P. Overall my recommendation was
generally for it but there was some caveats
around it that requires us actually coding why
we order stress tests which we currently do
not do.

DR. SPENCER: Kirk Spencer. I
really had almost identical gradings with
minor differences. I have a couple comments.
One, I would like to see the measure clarified
about whether the patients had symptoms at the
time of the PCI so the stress echo document
which makes and the stress nuclear document
make distinctions about asymptomatic angioplasty patients whether they had symptoms before their PCI or they did not.

If you have symptoms and had a PCI, if you are recent enough to come back you would likely get symptoms again so that is a very reliable thing to follow. Whereas, if you didn't have symptoms before your PCI and you are asymptomatic afterward, is that particularly reassuring. They don't make distinctions. They are talking about both groups of patients. I would like to see that clarified.

The other are minor issues. The second exception is they don't really deal specifically with is there is certainly wiggle room in a patient who has high-risk angioplasty looking for restenosis. In fact, the testing guideline makes it a 2(b) indication to detect restenosis in selected high-risk patients. We can either not exclude those patients and if you measure isn't zero
percent you are doing okay. You're going to have a 2 percent but it's a 2 percent that is defensible or we can try to exclude those patients. The guidelines in some respects haven't kept up to the angioplasty literature and the problem there is the left main. What to do with looking at restenosis in left main disease is sort of much less unclear. We are doing a lot more left main angioplasty. I think many cardiologists feel that it's appropriate to stress people to look for restenosis getting left main work done.

The third comment is there are certainly easier areas to pick off. When we talk about inappropriate stress testing, two of the other things we're going to talk about are the asymptomatic patients and the appropriateness guidelines and that got a 1. That's the lowest, 1 to 9. Instruct nuclear that got a 1. Everyone agreed.

The pre-op patients for low risk procedures both got 1 by both organizations.
PCI got kind of 3. If you're 4 you're indeterminate. You're a 3 you're inappropriate. The PCI area is not as clean as asymptomatic patients initial testing and pre-op in low risk procedures. This is a more gray areas and this doesn't come out as fairness to the gray areas of selected high-risk PCI patients and patients that didn't have symptoms before their initial angioplasty.

CO-CHAIR PETERSON: Comments from the group?

Ray.

DR. GIBBONS: Ray Gibbons. I guess I'm the champion of sample size issues in this group so I again want to raise sample size issue here pointing out that they are going to accept 35 cases and the four pilot centers that they had the rates range from 0.9 to 4.2. With a precision of .05 we simply aren't there at 45 cases and there even lower per the number of cases so the number of cases
really should be increased if this is going to be a reasonable measure.

CO-CHAIR GAZELLE: You are speaking to the denominator. Is it problematic to increase the denominator? You would think there are not that many sites that do fewer than 45 total.

DR. GIBBONS: It is number of PCI and the second division is the logistics of this, as Eric pointed out, a challenge. The more cases you have to do, the bigger the challenge.

CO-CHAIR PETERSON: I'm sorry, Ray. Forty-five cases was?

DR. GIBBONS: That would have been the CMS. They say 35 here.

CO-CHAIR PETERSON: Thirty-five.

DR. GIBBONS: So your precision would be less than .05 for 90 percent competence if you're zero.

CO-CHAIR PETERSON: I'm just thinking most are at the center level.
DR. GIBBONS: Right, center level.

Your precision -- I mean, I didn't do the calculation but it would be .04 or something like that for your 90 percent but that would take you from the bottom to the top on chance alone the next year in their pilot data. They didn't find large numbers.

CO-CHAIR PETERSON: The individual center does how many PCIs a year? If almost half those PCIs are done, even a quarter something means you have plenty of data.

DR. GIBBONS: I would agree with that but that's not what the pilot data shows.

DR. SPENCER: Oh, I sorry. This is Kirk Spencer. One of the papers, the feasibility paper looking at at least the nuclear half of this suggesting the multi-center declare appropriateness as one of the five most common reasons for an inappropriate nuclear stress test is in that table.

It says asymptomatic, post-revascularization less than two years after
PCI. Then it says symptoms before PCI so, again, in that document that is certainly a cleaner group. When you leave the asymptomatic group I think that is why the department gets kind of a 3, especially if we're thinking high risk. Some people are thinking they didn't have symptoms beforehand.

I think we could make it a 1 appropriateness if you make it symptoms before PCI and exclude -- I don't know how we define high risk intervention.

DR. GIBBONS: Ray Gibbons. Just a comment on that point. From our experience it's time to apply these criteria. Although clinicians commonly quote that issue, just as Kirk did, they actually go to medical records and see how well that's recorded.

It's recorded poorly for symptoms prior to PCI. At the time point you do this you would think most people would be able to tell you they did have something before their intervention within the last two years. They
don't do that reliantly.

CO-CHAIR PETERSON: Other comments?

MR. CORBRIDGE: We are moving faster than scheduled but he said he was going to be online at 1:15.

CO-CHAIR PETERSON: Okay.

DR. SPENCER: Well, I'll add another comment. Kirk Spencer. I would agree that although the symptom status at the time of the initial angioplasty clinically makes sense, it drops feasibility even further and we've already pushed it.

Because of the lack of electronic record we've already agreed this is probably a handwritten chart review sort of measure. Not only to you have to chart review now at the time of their PCI, or the time of their stress test, now you have to chart review two years back to the time of the PCI so I would be willing to trade off the symptoms for the following feasibility.
I think that's fair. Individually when I see the patient I can sit there and say, "Do you remember did you have symptoms two years ago?" That's very easy and appropriate to do but as a measure I would agree that makes feasibility difficult and probably should leave it out.

DR. GIBBONS: Does it impact the data that much to leave it out?

MR. BACKUS: Mike Backus. I don't have any asymptomatic patients getting PCI.

DR. SPENCER: Exactly.

DR. GIBBONS: Ray Gibbons. Because I'm going to have to go, let me just make more further comment on this issue. We have the experience in doing this. We would look at a clinician's note, let's say, in 2005 and they would say the patient was asymptomatic prior to their previous procedure. Because of our electronic record we could then go back and actually look at what the clinical note said before their prior
procedure and they would record the opposite. When my research nurses first came to me with this problem, it created an interesting methodologic issue which would we accept --

DR. SMITH-BINDMAN: What do you mean it was sort of the opposite?

DR. GIBBONS: Meaning the clinician at the time said they were symptomatic but the patient's recall was that they were asymptomatic or vice versa. We sometimes assume things and they will come up subsequently when Art discusses one of the other measures. We think when the patients report something that is actually what has happened but I can assure you as a clinical researcher, especially regarding stress tests and symptoms, that if you actually have the documentation to check on that you are surprised by how often their report of what test they had is totally wrong.

CO-CHAIR PETERSON: But I guess there is --
DR. RUCKER: My patients tell me different early inconsistent stories all the time.

CO-CHAIR PETERSON: We are going to have to get Joe's comment on how this would be operationalized. My general sense is at the time of a procedure somebody would code the indications for it. More broadly spoken in all of stress testing this has to happen. If they code it in theory, and we can talk about whether we would want this or not, you could code asymptomatic but was asymptomatic prior to or something close to that. You could put that in there. Or you could choose not to and just say asymptomatic and we'll just say there is a relatively small collection of patients.

DR. SPENCER: Kirk Spencer. We're trying to get rid of the ones that discuss post-PTPA which is what many of them are now.

CO-CHAIR PETERSON: So Joe?

DR. ALLEN: Yes.
CO-CHAIR PETERSON: Welcome.

DR. ALLEN: We did not choose to include symptom status in this measure prior to the PCI even though some of the original corporate use did have that as a caveat because of the feasibility issue that Kirk had been talking about as knowing what the symptom status was as long as two years ago.

CO-CHAIR PETERSON: Joe, you can say whatever you'd like but one thing is if you could just give a little bit of a background of how you believe this would operationalize out. What would happen moving forward to allow this to be feasible?

DR. ALLEN: Sure. Probably many of these tests have been reviewed in a very inefficient way which is the third-party review. It is happening even if this measure doesn't go forward before collecting information.

The measure is meant to put some parameters around what it is that is meant by
inappropriate measuring. We also have in ATC a number of mechanisms in a number of facilities that have instituted electronic data collection of this type of information. We have both lead based and registry based ways to collect this as well as decisions in talking with vendors about implementing their decisions.

Although like some of the measures that were discussed yesterday, it could be that a lot of places might not have the capability to do it. Right now they are doing it very inefficiently by calling the third party. We believe that in the very near term there are electronic commissions to do this and most actually prefer the electronic data collection that right now requires a phone call.

CO-CHAIR PETERSON: Just to clarify one issue that has come up on the committee in the space of transparency. A couple of issues. One, while the ACC has
developed these criteria, the criteria in the public domain, you may or may not be developing a product in-house and/or with other vendors but there are other vendors who would be able to develop this product who are developing these products rapidly as stand-alone. Correct?

DR. ALLEN: Correct. We want the measure to be out there based on the criteria, transparent and anybody could use it. We would develop a program around it for full disclosure but we believe the way it's implemented that we are going to have it out there and the product advantage would be the program, not the measures or the --

CO-CHAIR PETERSON: Great. Are there any other statements that you have for us? We can ask you some more questions.

DR. ALLEN: Actually, I have some more questions. I know something came up yesterday and I've heard the full discussion today so let me see what questions you might
have and then I can ask others.

CO-CHAIR PETERSON: Just a second.

Kirk, do you want to ask a question?

DR. SPENCER: I was getting Dr. Gibbons' comments on the other ones we're going to discuss before he leaves.

CO-CHAIR PETERSON: Okay. He's ready for your question.

DR. SPENCER: So I think we agree on the chest pain two years ago making the feasibility too low. Did you discuss -- one of my senses is that appropriateness criteria both rated these as 3's and not 1's, high-risk angioplasty.

What do you do with the selected high-risk patients or the left main patients. We just leave them in there and then if you have a rate of stress testing that is 3 percent that accounts for those, that's okay? It's a little cleaner that the measure should be kind of zero. Was there discussion about that?
DR. ALLEN: Both on the appropriate use criterion. In developing a measure there is always the -- which may be more precise and how much data collection and feasibility you get into so we don't believe that these rates will go to zero. We know that based on our pilot study that they are as high as 15, 20, or 30 percent overall for inappropriate. The things we are focusing on in these measures are the top three issues. You are correct that we would assume that those cases that would be exceptions would be in the ones that you would show as your 3 or 5 percent, whatever it ends up being, kind of the low rate after you view the things to get rid of the patients that really shouldn't happen.

DR. SPENCER: So you think it's too hard to pull out selected high-risk patients with left main angioplasty, proximal IB angioplasty, that there be some agreement on betting asymptomatic stress may be
appropriate?

DR. ALLEN: We didn't rate that for the reason that we just talked about which is the additional data collection we didn't feel would be feasible to go to that level. If we find that we continually have the 3 to 6 percent and it seems like it's coming around to similar issues, we could always revisit that but we feel like this was a reasonable starting place.

DR. SPENCER: I guess the intended harm there, then, is if you're a center high-risk angioplasties often get sent to specific centers that your inappropriate rate will be higher if you do high-risk angioplasty so we just have to list that as an unintended consequence, I guess.

DR. ALLEN: Right.

CO-CHAIR PETERSON: Joe, another question. Any reason you didn't include bypass surgery, asymptomatic bypass surgery?

DR. ALLEN: We discussed the
bypass surgery as a part of this measure. We felt that, given the different time frames, that we did not want to do that for this particular measure and it didn't come up as frequently in the pilot either as one of the reasons so it was both based on pilot data and different time frames.

We didn't want to send the message with the measure that, if we put it at two years to make it equivalent to PCI, that it might send the wrong message and then just from the pilot data it didn't seem to raise up to the common type issue.

DR. SPENCER: The pilot data did include symptom status at the time of their initial angioplasty. Right?

DR. ALLEN: Yes, it did.

DR. SPENCER: So it wasn't that hard to get.

DR. ALLEN: We found that we got a lot of questions on it and we didn't know if the reliability was as good as many people
recorded it and whether or not it was reliable.

DR. SPENCER: Do you have non-published data to suggest that it's different or better than the patients that didn't have symptoms as thought they recorded as it could be recorded?

DR. ALLEN: No, we don't have data that would say there is difference.

DR. SPENCER: Subtract the total patients from the ones that had symptoms. You've got it but you haven't analyzed it, I guess. If you know who was symptomatic, you know who was asymptomatic. You just don't know the number. Okay.

DR. D'ORSI: Excuse me. Is it ever valid to have one of these stress tests with asymptomatic patients after two years of a PCI?

DR. ALLEN: So, after. You're saying not based on this measure but after two years, would it be something that you might
reasonably do.

DR. D'ORSI: What I'm saying is, if this measure comes out three percent and this is a quality forum and this is an outcome measure, how would you interpret that.

DR. ALLEN: There are three percent that are getting it. Even though we have said it's inappropriate, there are reasons that they might have received it and could we look at that based on the data why they might have received it different than outcome?

DR. D'ORSI: Yes. In other words, I'm getting at why do a metric if we don't know what to do with that number.

DR. ALLEN: What all these measures focused on is where we don't have a demonstration of benefits based both on risk and on other factors that would show that these patients should be getting stress imaging so it's not -- unlike an under-used measure which ties to and improves outcome per
se, these are more clearly efficiency measures where you are using a resource and putting the patient through potential downstream impacts. You could look at, if you did want to, and I'm not sure you would get something different, just whether or not patients avoiding this didn't have subsequent procedures or something like that that might be a temptation once you start down the stream of seeing something and starting in on the pilot data.

You see something and then you follow up with a CAS or a CTA and things like that so if you could look at a briefer compensity, you know, in reducing this based on this measure.

DR. D'ORSI: Carl D'Orsi again. So do you feel perhaps it's a little premature to make a measure, a quality measure, out of this without a little more data?

DR. ALLEN: I'm sorry. Could you repeat that?

DR. D'ORSI: Yes. Do you think
it's a little premature to make a quality measure with an outcome end point at this stage?

CO-CHAIR PETERSON: Can you clarify your question?

DR. D'ORSI: In other words, we are making a quality metric with an import. We're saying, first of all, we don't know what's good or bad. Your paper addresses how many are getting this test and CABG.

CO-CHAIR PETERSON: We do. We do. What the agency is saying is they do. They are saying that in general that in this indication all asymptomatic patients after percutaneous intervention testing is probably inappropriate. Then you stretched it and said, can you come up with any indications in a patient somewhere that would fit it? His answer was, well, yes, maybe. So should the number absolutely fall to zero? Maybe not but, again, this gets back to Helen's point that NCDR, in this case National Quality
Forum, is not in the business of trying to find every single exclusion that might exist on the planet but rather to get --

DR. D'ORSI: Right. Well, okay.

CO-CHAIR PETERSON: Because it's interpretable and then interpretation is, yes, everybody in the world has it down at two percent and your site is at 40 percent.

DR. D'ORSI: Okay. I understand that. Could you at least get a range that is acceptable then?

CO-CHAIR PETERSON: It should be close to zero.

DR. D'ORSI: So zero to one is acceptable?

CO-CHAIR PETERSON: Unless you do a lot of high-risk angioplasty.

DR. D'ORSI: All right. Thank you.

CO-CHAIR GAZELLE: This is similar to the discussion we had about age-stratifying of mammo measures. The question comes down
to, are we willing to accept that there is
gothing to be some range, some variation that we
could explain versus we want to narrow it down
to no variation. As I did yesterday, I would
lean toward allowing there to be some
variation and having fewer exclusions and then
just understanding that we are not drawing a
threshold anywhere.

DR. SPENCER: The only problem --
Kirk Spencer. The only problem with that is,
again, it's a public measure. Patients who
get MRSA in the operating room we understand
that. It should be zero. That's really
clean. There is no good reason to get an MRSA
in the operating room.

Public measures that you should be
zero and you're not zero as a public measure -
- we understand that as doctors, that 3
percent is probably right.

DR. SMITH-BINDMAN: Is there a
need for -- this is Rebecca Smith-Bindman. Is
there a need for face validity to include
those exclusions? You just have to just help us understand the magnitude of this.

CO-CHAIR PETERSON: Right. Why don't you ask the question with regards to the high-risk angioplasty.

DR. SMITH-BINDMAN: How many are there?

DR. SPENCER: I bet it's small. In a high-risk center it's probably still three percent, five percent. And where that should be even stressed there is even contradictions about that. Therefore, it wouldn't be a three. Their score would kind of a six.

CO-CHAIR PETERSON: We are nervous about doing a procedure which we -- we are still nervous about doing it.

DR. SMITH-BINDMAN: There is still face validity in this measure without having all this.

DR. GRIFFEY: Just because Ray is not here -- Richard Griffey -- do we still run
into the same issues that he raised concerns about with respect to small sample sizes and small facilities?

CO-CHAIR PETERSON: Yes, Joe. The issue was raised as to how many cases most centers saw that were done under this indication.

DR. ALLEN: Say that again? How many cases?

CO-CHAIR PETERSON: At a given center how many cases given Ray's question of whether or not we had a sufficient sample size at any center? How many patients would get testing for this reason?

DR. ALLEN: Right. We looked at that in our pilot data and the reason why we chose the 60-day time frame as to put this at the imaging lab level to get enough volume for each one of these measures that have to go back into the actual data for a pilot. We did look at that and made sure that the majority of groups, and we had different sized groups
participate in our pilot, that all them could
collect enough data in 60 days to have at
least 30 cases.

CO-CHAIR PETERSON: So you are
saying within 60 days you would have 30 cases
that would fit this indication? I don't want
to pin you down but --

DR. ALLEN: Right, right, right.
We looked at whether or not we could collect
the information on the number for the
denominator that there would be at least 30
cases. Whether or not there would be 30 PCI
cases we looked at the -- we could find some
cases in those 30 that were PCI, and most
centers did, but it wouldn't be 30 PCI cases.
It would be 30 for the denominator.

CO-CHAIR PETERSON: Okay. Do you
have any idea how many cases might happen in
a typical practice in a year currently that
would fall into this category?

DR. ALLEN: I don't have that off
hand. I'll have to look it up.
DR. SPENCER: This is Kirk Spencer. Is there any reason you would mind including stress MRI and CTA that is also not appropriate within two years? We don't want unintended consequences to drive all the business to CTA.

DR. ALLEN: Right.

CO-CHAIR PETERSON: And the second rationale for that is actually we have another measure pending that looks at a broader set of stress testing and I can't think why we wouldn't do that.

DR. ALLEN: That is a reasonable suggestion. The reason we didn't include it this last time was because we were still updating the PT document that didn't speak to that. It does not speak to that and we do have criteria now on that so we can update it.

CO-CHAIR PETERSON: Okay. Does anybody else have an issue if we agree to that as a conditional amendment? No negatives? I'll take that as a no. Okay. Any other
questions for Joe?

DR. GRIFFEY: I have a question.

This may be more for you all. Do you find sufficient reason that we would not try to combine this measure with the one we previously discussed?

CO-CHAIR PETERSON: The data are different so we would have to say extend their measure to include post-CABG. Is that what you're proposing?

DR. GRIFFEY: Yes.

CO-CHAIR PETERSON: The question is a revamping of a question I asked earlier to you. You may not be able to answer this today or not. If not, we can come back as a conditional and then we can come back with a response one way or the other but re-raising the issue of including expanding this to include bypass surgery.

DR. ALLEN: You know, I think, as I said, being that we could look at -- the rates were kind of low for a look at another
type of procedure and, again, you know, the
different time frames. We would have to
reframe the measure.

CO-CHAIR PETERSON: But the
measure is more extreme. Right? So if it's
testing within the first two years when
testing within the first five years was
inappropriate, that seems to be okay.

DR. ALLEN: Right. It would just
be additional data collection to look at the
additional patient population.

CO-CHAIR PETERSON: Can we be
clear? One last thing, I'm not going to push
you here. The way you're setting this up you
would basically -- for most centers wouldn't
they need to code most of their indications
for procedures to be able to collect this?

DR. ALLEN: They'll have to at
least look to see if they finished it based on
that first measure and to any one of these
categories of did they have a PCI, were they
asymptomatic or their pre-op testing.
They don't have to do any further coding of the patients once they have packed them into the two categories. The patient then qualifies or any one of those three things and those that don't. By adding CABG you add to an important category that they would add a few more pieces that they have to evaluate.

MR. BACKUS: This is Mike Backus. So are we expecting for data collection that this is done post-service chart review or are we expecting that this is done pre-service filling out a form which is I know how some of the Brighams --

DR. ALLEN: This is measured at the laboratory level so it is at the point at which the imaging delivered. It wouldn't necessarily be at the point of order and so it's like sort of the same. Maybe the measure would be aggregated at the lab level.

MR. BACKUS: No, I understand that. My question is is the data collection
after the image is done on a chart review? We did some stuff for Brigham and Womens this morning.

DR. ALLEN: We are just matching prospective data collection efforts because of issues that -- you know, finding some of these things in retrospect.

MR. BACKUS: What we're expecting, just so I understand it, we're expecting a lab or a physician to fill this out and say, I'm doing this pre-service, and we're expecting them to say that essentially I'm going against the ACC guideline at the time they fill out the form and then submit it?

DR. ALLEN: The question here is why would anybody ever submit an inappropriate order. What we learned from the pilot data was the rate of orders per individual physicians as they come in, they are coming in from a number of different places because that's one thing. Usually you have an average of five imaging tasks coming from any
individual physician into the imaging lab so a physician orders five over the course of a year.

On average even a cardiologist only 30 over the course of a year. On an individual case basis a lot of physicians will submit inappropriate orders because they feel, well, this particular patient I can think of reasons why. I can justify in my own head.

MR. BACKUS: Right.

DR. ALLEN: What we find is once we get the pattern back through the measure that they see that, oh, my one contributed to a rate of 15 percent overall for inappropriate imaging. Then you start to educate that about, okay, even though you can individually justify can you rethink about these particular three or four issues when you go to order so that you can reduce that number?

You start out usually with this inclination to prompt the order and even do an event inappropriate and then when it's
aggregated at the lab level you get enough cases and feedback and you see how that one contributed to a larger issue.

CO-CHAIR PETERSON: Point of clarity.
We actually yesterday passed the same measure but you would have to put in for the study based on the Brigham system as inappropriate and circle that inappropriate back. We did, in fact, by definition drive it down because we do that.

DR. GRIFFEY: Well, in fairness you had to specify your indication. That's what you had to do.

CO-CHAIR PETERSON: That's what we're going to get into.

DR. GRIFFEY: I understand that but, I mean, saying that you're going to enter an inappropriate indication up front makes it sound kind of silly whereas at that point the person may get some decision support from that and decide against doing the study, or they may have a reason that they feel is
appropriate and they may indicate that reason
at that point.

I mean, someone is not obviously
going to enter something knowingly this is not
recommended and intentionally writing this
like we said and so that is the value from
that kind of decision support up front. And
the issue is feasibility.

MS. ZERZAN: But I think the
difference is in those measures that person
came to the ED with a complaint. People
aren't coming to the cath lab saying cath me.
They have an appointment that is made for
that. It's kind of a different clinical
situation and I think there would be a ton of
pressure to go along with the procedure once
you have it scheduled and a patient is there
expecting a cath.

DR. SPENCER: Agreed.

CO-CHAIR PETERSON: So, Joe, this
gets back to the issue that -- you're
proposing that for all tests that are ordered
whether point of order or point of service, hopefully point of order, that you would be getting an indication for that test so that would be for all tests.

DR. ALLEN: Okay.

CO-CHAIR PETERSON: So that would be for all tests.

DR. ALLEN: Correct.

CO-CHAIR PETERSON: At that point we're done. Right? Because we would have enough information because you would be collecting information at the point of that order.

DR. ALLEN: We would certainly know the date of their angioplasty.

CO-CHAIR PETERSON: You're proposing to collect that all at one shot, not to have it required beyond that, or are you?

DR. ALLEN: No, we would collect it all up front and then feedback the pattern over time. We both get, as you said, the upfront. You know you are being watched for
these things and you are going to avoid them.

If you are inclined to say, well, but my
patients are different and special each and
every time, and that aggregates to a pattern
for an imaging lab, then they can pick that up
and do some education with that particular
group as to why this was deemed inappropriate.

CO-CHAIR PETERSON: Any other
questions for Joe? Any other comments from
the committee?

DR. RUCKER: Can you speak sort of
one more time to the end number? I'm not sure
from the other comments. What percentage of
the sites do we have a large enough sample
size to do this? I would be happy with just
the gut feel.

CO-CHAIR PETERSON: Out of a
typical center we anticipate how many centers
we anticipate will have more than 20 bases
that would meet this criteria.

DR. ALLEN: The majority of
centers because we are aggregating at the
imaging lab level would have enough volume in
the window that we are asking for this measure
to get data back. We have looked at their
pattern both at 60 days and at one year and
they didn't change based on how many patients
came in.

DR. RUCKER: Okay. Thanks.

CO-CHAIR PETERSON: Any other
questions? Comments from the public? So we go
to voting.

DR. GRIFFEY: We decided there is
no entertainment of combining measures.

CO-CHAIR PETERSON: I think it's
up to the committee now. We had one minor
expansion. This has already been conditional
and everybody agreed on the idea of
conditionally expanding it to include the
other CTA and MRI. That has been agreed to.

The question of the CABG can be
our committee's decision if we choose to do
that. Let's vote on that first and then vote
on the whole thing. I think that's fine.
DR. GRIFFEY: I was just thinking if there is an issue of low number and you made it one or the other, then you just increased by some percent. If the criterion standard is stricter --

CO-CHAIR PETERSON: We can put that on, they can come back, agree or not, and we can pass it conditional or not. If we want in general to do that, if we were going to pass it, we can vote for it now. How many are in favor of increasing it to the PCI and CABG? Okay.

Now it will be conditional on two conditions both expanding the test and doing --

Any other conditional changes? Okay. Moving onto the importance of the question. How many feel it is high importance? Moderate? Low? Okay.

Scientific acceptability? How many vote this high? How many vote this moderate? How many vote it low? Okay. Usability. How many vote
it high? How many would vote it moderate?

How many would vote it low? One solid low.

Okay now, feasibility. How many are voting high? How many are voting moderate? How many vote low?

MR. CORBRIDGE: Could we do the low one more time?

CO-CHAIR PETERSON: Sure. Low? Anybody else want low? The final is to vote on the measure. How many want to see this measure passed with these two conditions? Their arms are going up and down.

MR. CORBRIDGE: One more time. Okay.

CO-CHAIR PETERSON: Joe, are you able to stay on the line with us for awhile?

DR. ALLEN: Yes, I can stay on for a bit.

CO-CHAIR PETERSON: Good. Okay.

So we are going to go back to further discussion on the CMS measure for bypass surgery. Further discussion now that we've
had this discussion hopefully. The issues that were before us before had to do with issues of the inclusion criteria were too broad or were not broad enough to capture an asymptomatic population. That was agreed to. Everybody agreed that we would add outpatients.

DR. RUCKER: Hadn't we just added CABG to the --

CO-CHAIR PETERSON: We added CABG to a measure that was done through this --

DR. BURSTIN: NQF will endorse measures based on different data sources if it's appropriate and adds value.

CO-CHAIR GAZELLE: Now we are considering number 11.

DR. BURSTIN: Number 11. Admin. data over age 65 CABG only.

CO-CHAIR GAZELLE: Other discussion?

DR. SPENCER: So does this add value if we have the ACC data? I think this
is a feasibility issue. We have every center in the country whether they have decided to fill out the ACC forms or not. That's the value getting at the same idea.

MR. BACKUS: This is Mike Backus. I think what you're going to get is right away some validation or not of how well the coding works. Either the two data sets are going to converge and that is all great or they are going to diverge and that's not bad because it will fortunately get closer to the heart of the question.

DR. SPENCER: I'm sorry. Is this time limited?

DR. BURSTIN: The CMS 1 is tested so it's not time limited, I believe this one is.

DR. SPENCER: I will make a proposal that I had some discussion with CMS during lunch and they are willing to look at changing the ICD-9 criteria and rediscussion of the catheterization number 2 criteria. It
sounds like the third criteria really won't be a big issue.

We give them time to rewrite and revote but I'm letting it die this -- it's important enough to not let it die this cycle because we thought the imaging thing is not coming up for two more years. I guess we can turn it around in three or four weeks.

DR. SNOW: Just leave it tabled then.

DR. SPENCER: If they let it die, then it dies. If it comes back, we revote.

DR. DEHN: If I could speak from a developer's standpoint, I mentioned before we have a smaller group in Indiana and the result of that was a long list. It sounds like this group wants a shorter list. We can accommodate that without difficulty but would like some direction from you.

We can certainly turn this around in no time if you would authorize some of the cardiologists on this to work with us we can
do that. I just don't want us to go through developing a whole list and then it still doesn't meet what this group wants and this is the group that really counts.

CO-CHAIR PETERSON: Just for point of clarification, we had an offline conversation while you guys were talking. The last measure that we have now approved we are going to approve with time-limited even though there is some degree of testing that has been out there. The degree to which that has been spread beyond a sort of pilot test effort we'll probably need some data.

Joe, we'll talk to you offline but the ACC will have to develop a plan for how they would implement this and get further testing of its applicability.

DR. BURSTIN: It looks like there is data for liability.

DR. SPENCER: Can you make a note that was done after the vote?

CO-CHAIR PETERSON: It is a single
modality and no CABG data. CABG has not been included.

DR. FIESINGER: Other question. This is Troy Fiesinger. On the proposal to reformulate the CMS measure are we going to add post-PCI to that parallel with the ACC measure? There is one about changing the exclusions. I would like it to be parallel if we're going to try to compare, as Michael pointed out, to CMS database data.

DR. SPENCER: This is Kirk. I voted against combining them and the ACC did. There are different procedures. The reasons to do a stress in the two I can imagine different scenarios. They are very different.

CO-CHAIR PETERSON: I respectfully disagree. Is there any other discussion?

MR. BACKUS: This is Mike. I think the thing would be if you could stratify the CMS data between the two because the question to me on the ACC data is not the measure or the value of the measure.
It's the feasibility of the measure and so to the degree that the two datasets show the same thing and one has virtually no cost to the practice whereas the other is a manual question or something to be developed, then if we see the two shake out the same and the group becomes happy with the asymptomatic issue, we've got the same kind of measure, same dataset, and we've taken burden out of the practice.

DR. BURSTIN: Although there would still be additional work to make a CMS measure work for the under-65. A research recommendation would be, I think, going forward you would also want to be able to get the admin data to look at what they're doing.

MR. BACKUS: The imaging rates in over 65 is so much higher. It's three to four times.

CO-CHAIR PETERSON: Any more comments on this proposal on the table which is to have the measure expanded to include PCI.
but to report it as a CABG and a PCI measure separately?

DR. BURSTIN: That is acceptable to the measure developer to consider.

CO-CHAIR PETERSON: So that we would expand this to include PCI but we would report out CABG and PCI separately.

DR. SPENCER: The two-year PCI and not the five-year like the CABG?

CO-CHAIR PETERSON: Yes.

DR. SPENCER: On paper CTA is not an issue. Right?

DR. DEHN: If you don't need a stress test it's fine. We threw it in the ACC measure.

DR. BURSTIN: The bottom line is that we should come up with a harmonized measure to look at.

DR. RUCKER: Don Rucker. CTA, I think, also has a different behavior. If you are doing it for people who have known coronary disease which presumably is the case
for people who have had a CABG or PCI, in

terms of durability finding plaque, I mean,
we're talking that we're moving from
luminology kind of studies to studies that
actually show intrinsic wall disease.

In the sort of people who are not
known to have heart disease, it's a very, very
different performance and potentially very
different durability. Here I think it's not
as important honestly because you already have
known disease and you are really looking at
luminal issues more than you are presence of
disease but I would just throw that out.

CO-CHAIR PETERSON: I think you
wanted to get some guidance from our
cardiologist here.

DR. RUCKER: The other aspect is
that for a long time these were local coverage
decisions and they are not all the same so
that there is coverage in some areas of the
country and some there isn't and that sort of
differentiation and variation we thought --
CO-CHAIR PETERSON: Neither point, I think, would change your proposal including this.

DR. RUCKER: They would argue that it may or may not be covered but who cares.

DR. BURSTIN: I would suggest as you vote on this we'll vote with the expectation that we are going to get a revised measure back and take a closer look at it, obviously, after discussions to make sure it actually meets --

CO-CHAIR PETERSON: Can we vote on a revised measure with this many revisions?

DR. BURSTIN: It's up to you.

CO-CHAIR PETERSON: We should agree to delegate the cardiologist leads out of this group that have devoted the time already to coordinate and then come back with something that is consistent that they can recommend.

MR. BACKUS: I'm willing to do that. I would like to be with some of the
people from the original group so we're not missing something they were thinking.

DR. FIESINGER: I would not like only the cardiologists involved. I want to be very clear on this. One of the challenges, if we take all the exclusions out it becomes, I personally believe, a non-existent -- we are going to be so much pushback from all the cardiology community that it could be invalid by the community to go to the numbers. I think we need to get other people on the table that can look at that as well.

CO-CHAIR PETERSON: I think it will be up to you all to pull what parts of the committee you choose to. You can choose to use us as resources or not. Ultimately this will come back to this committee for a revote. If we want to take a straw poll now to say would we be interested in revisiting it, I think in general the feeling is we would be interested in revisiting it, but no promises. Peter or Janice or somebody, be sure to send
me your paper.

DR. BRUETMAN: Sorry if I'm not
catching the whole picture but we've heard a
lot of things that were asked and I want to be
sure what the expectations are because it's
going to be a significant level of effort for
us.

I mean, five years of data and,
oh, we also wanted this and that. It won't
turn out like discussions on exclusions and
rewriting the data cost a few hundred thousand
dollars to do it again and I don't want you to
say, we should have done this.

DR. BURSTIN: If you all would
write up exactly what the committee based on
the discussions today exactly what the
committee is requesting, we'll run it back by
the committee and then we'll send it to you so
you don't have to do anything until we've
gotten agreement from the committee that
that's what they want.

CO-CHAIR PETERSON: Okay. Moving
on to 17.

DR. SPENCER: I think we can do this quickly. So the idea on 17 is looking for inappropriate people to stress in a group. The three kinds of people we are going to look at in separate measures are initial assessment of asymptomatic patients, routine testing after PCI, and preoperative testing in low risk surgery patients but it's not looking at the rate of stress in those three groups because those are all three separate measures.

It actually looks at the proportion of test requisitions and the patient's chart that documents the use of a nuclear stress echo with adequate data to demonstrate avoidance of the common inappropriate uses. It's kind of a very different measure.

Of people that have stress disorder how many of the charts have enough data in them to prove it wasn't for all these three bad reasons. I would propose that maybe
-- well, you don't know enough to vote no
against it yet. The three measures are
already identified so we are already looking
at asymptomatic patients, we're already
looking at post-PCI patients --

MR. BACKUS: I don't have a 17. I
have two versions of 16.

DR. SPENCER: Me too.

CO-CHAIR PETERSON: I have 17. It
was in the documents they sent out last
Thursday. For some reason the bookmark is
wrong.

MR. BACKUS: Oh, the bookmark is
wrong, that's all.

CO-CHAIR PETERSON: The bookmark
goes to 16. Just scroll down past 16.

MR. BACKUS: Oh, so it's like a
type 1 or something.

DR. SPENCER: I think two reasons
why maybe we don't need this one is, again,
all three criteria are separately identified
as looking for overuse. The other one is it's
a bit of a difficult thing. If the numerator
is the number of charts that have data to
answer but the denominator, again, is already
patients that are post-PCI, pre-op, or risk
stratified or asymptomatic so, in some
respects if you already know the denominator
is a post-PCI patient, you sort of already
know in the numerator that it was a post-PCI
patient. That is why the test was ordered.
I mean, that's why the test was done. You
don't know whether it done in under two years
or the chart doesn't document that but it's a
measure I don't think we need and it's a
measure that I don't think the numerator or
the denominator are different from each other.
I don't know if that made any sense. It's a
funny measure. How many charts of patients
that got stresses for one of those three
reasons have their data to tell me that is the
reason why they had the test.

DR. GRIFFEY: It would have been
nice for just an indication for example.
DR. SPENCER: What do you mean? Do I have enough data to prove that it was or wasn't ordered for one of these three bad reasons.

CO-CHAIR PETERSON: So you're saying it fails on the importance.

DR. SPENCER: Yes. I'm suggesting we go with unimportance.

CO-CHAIR PETERSON: Joe, you want to have any comments? The issues have been raised about importance. When you fail on importance you fail on the measure. Do you want to address that issue?

DR. ALLEN: Sure. We developed this measure to avoid one aspect. We know that as far as people can record different pieces of information. I thought the easiest way to gain all their measures is to just vow to document anything in the chart related to these issues and, therefore, the patient comes up uncategorizable rather than being able to assess them into one of these categories.
It is the point that was brought out related to having the dates recorded, having these different pieces of information that we need for each of the three recorded, things like that, so we can evaluate this. We understand there are some concerns about importance.

We feel that it will help avoid people just not recording things and then, therefore, doing better or looking more like they aren't doing inappropriate cuts just because they fail to record that data.

CO-CHAIR GAZELLE: This is Scott. I don't think you need this at all, I agree, because it's already implied if we dump any of the other measures you've got to have accurate data so to have a free-standing measure that assess the accuracy of data without having a measure that assesses the use of the exam wouldn't seem to be appropriate for this form in my opinion.

DR. SPENCER: There are also a lot
of legal requirements that we have to have this data.

CO-CHAIR PETERSON: Okay. So we'll vote on importance. Public comment?

DR. BURSTIN: No.


Moving onto -- anyone have a preference? Let's do the other ACC measure, 16.

DR. STILLMAN: This measure is cardiac stress imaging does not meet appropriate use criteria: Testing in asymptomatic, low risk patients. We talked before about how we believe this is an important area and essential to improve efficiency.

I have a list here. So in terms of demonstrating high impact on healthcare and citations, I gave that a C. Certainly the paper supports this.
Opportunities for improvement, I also gave it a C. There are a substantial number of patients who are asymptomatic with variability that could be improved upon.

For the outcome or evidence to support the focus I think there is adequate evidence. I gave that a C.

Moving down to 2, the numerators, this is the number of stress SPEC images and stress echo performed for asymptomatic low CHD risk patients for initial detection and risk assessment. This was done with a number of exclusions and I think there are some issues with this. The first issue is low risk because the way the risk is being assessed is by the clinician. It's an opinion. What would be more appropriate here, I think, would be an objective measure such as a priming at risk

Ray Gibbons when he was here earlier commented about a study that he's aware of which risk assessed by a physician
versus an objective measure could vary quite a bit -- so I think that in itself could be a problem.

The exclusion criteria I think had the benefit of providing a more uniform sample but there are a number of issues, I think, that are related to it. It's not always clear. Patients aren't always certain about what test they had so you may not get good data here to begin with. That was another comment that Ray had. I think there are unintended consequences from the exclusion criteria which we will discuss later so I gave this a P.

The denominator is the number of stress SPEC MPIs and stress echoes performed so it's pretty straightforward.

Moving to reliability testing.

CO-CHAIR GAZELLE: May I just make some clarifications? The denominator is an exclusion that I didn't understand. It says, patients without collection criteria recorded.
Isn't that kind of a squishy easy out?

DR. STILLMAN: I think that's a good point. We could be explicit as we found out.

CO-CHAIR GAZELLE: Right but might not be excluding many of the patients who are trying to identify it?

DR. STILLMAN: That's a good point. The intent here is to use registry data or RAV data. The reliability testing, again I think there is reasonable support for this so I gave that a C. For validity I also gave that a C also because there is reasonable support.

The evidence supporting the exclusion criteria, I think the intention here is to be certain that the patient doesn't have known coronary artery disease so I gave that a C. The risk adjustment outcomes, the resource measures are given and there was no risk assessment.

For identification of differences
in performance, C. Comparability of local data sources and methods, N. There really hasn't been much done. NA for disparities. Usability I gave that a C. Harmonization, NA. Distinctive or additive value, a C. For usability I gave that an overall C. Feasibility. The data generated is a byproduct of care processes. I also it a C. For electronic sources a P. The reason why I gave that a P is, again, it's not always clear in electronic records what procedures patients have had. It might have been done at other facilities. It's going to be a bit of a dirty dataset. For exclusions I gave that a P. Again, the importance here is to have a uniform dataset. I mentioned earlier the problem of having some unreliable exclusion data but I think worse than that is going to be the unintended consequences for it. Those could be, for example, you
want to do a SPEC study you can't do it if the patient has had a calcium score so you might be inclined to do a calcium score or do a CTA or some other test in order to be able to do your SPEC study. I think it really has a risk of driving up other testing.

In the end for feasibility I gave it an M and my final recommendation was not to approve this as written.

CO-CHAIR PETERSON: Okay.

Comments from others in the group?

DR. SNOW: Yes. Interesting to me, it came out very similar in some ways but generally more harsh. I hate this measure.

CO-CHAIR PETERSON: Don't mince words.

DR. SNOW: I think there is an issue here. In fact, one of the things --

CO-CHAIR PETERSON: So just to be clear when you vote, H does not stand for hate.

DR. SNOW: This is Snow by the way
for the record. A lot of my displeasure was
in the second item because I think it's
already mentioned but I'm much more worried
about it, the risk being done by clinician estimate.

Ray was particularly eloquent in
that particular one because they did a study,
the Mayo they can do it much more easily than
in other places, in which they had clinicians
estimate the risk and then went back and
looked at the data and they found that the
clinicians had overestimated the risk very
substantially and consistently toward
performing with procedure. I think there is
an inherent moral hazard in the way this thing
is structured that's very hard to get out of.
The two really had me going.

There were also some issues around
the volume. Again, 30 cases is a low number
and you are going to really penalize a smaller
establishment. At the end of it, there were
several instances in which Art gave something
a B or a C and I would give it an M but we
came out no for the global assessment I think
for similar kinds of reasons.

DR. STILLMAN: Can I add
something? I think some of the issues here
can be repaired potentially so for the
exclusions if it could be maybe just changed
to no known coronary disease, no history, I
think that would take care of a number of
issues.

For the risk assessment if that
were made a quantitative objective measure
like Framingham I think that would address
issues. Whether that could be done within the
registry and the labs being submitted I don't
know. Perhaps Joe can address that.

CO-CHAIR PETERSON: I think Joe
will address this. My suspicion is that the
tradeoff here -- this is where we really have
to explain the judgment here. The tradeoff is
just how difficult it is to calculate these
things and how often they are calculated in
the real world which is almost none.

The flip of that is to say which
way does the bias go here? Do physicians over
or under code risk? We tend to overestimate.

DR. SNOW: Overestimate.

DR. GRIFFEY: Except it's all
retrospective data.

CO-CHAIR PETERSON: Right. But
even a stronger measure do it public or report
it and we will really overcode risk. I guess
the point being this. The only defense I
might come back to is to say if you actually
did code it and it's low risk and/or it's
potentially eligible for chart reviews, you
could get around this by not doing it and then
if no one met the low risk category you could
do chart reviews or something.

I don't know. Joe, I'm sorry. I
shouldn't speak to this. Joe, I think it's
perfectly reasonable to have you speak to this
at this time.

DR. ALLEN: We had an extensive
discussion of risk calculation when we were developing this measure. You'll notice in the specifications that we do say that we would encourage folks and would require folks in calculating their risk to use all available variables that they had for Framingham meaning that everything that is available to them at the time when they are doing this they should use that to calculate risk and to use age-based and gender-based averages for those that are missing. The most common ones we found in the pilot were cholesterol values that were missing. Our group in the substitution is a month between a full risk calculation meaning that the measure would actually collect all the variables and then do the risk calculation to verify down the other side and just having the physician code for the measure the risks that they calculate based on what is available at the time when they are going out to do some information.

We found the data collection a
burden to have folks submit every variable for Framingham and then on the backend ensure that was calculated properly for measure data collection was both too high on the clinician because they may be missing a couple of variables, specifically cholesterol values, and the folks that would actually be calculating it that they would have many more calculations but that we would do two things. We would require them to use as much information as was available, and then, as CO-CHAIR PETERSON said, if we were to go back we would say of all the variables that are available in the chart was it really used and do an audit.

Those are the two approaches and we would be open if there are suggestions that people feel like we need something more objective that goes further but we did do the tradeoff. In the pilot many times cholesterol values weren't available and it took much more additional time to do that and there was a lot
of calculation burden on the measure side actually because of each one of these.

CO-CHAIR GAZELLE: Scott Gazelle.

I'm leaning towards Roger on this. I think this is a terrible measure. I think it's a terrible measure because, first, I do feel we need a measure that gets at appropriate use of MPI. Everyone agrees it's growing fast but I don't think this gets at it.

I've been staring for the last 10 minutes at the numerator and I can't understand it. I'm sorry. Maybe it's because it's late but what I can't understand so the numerator is basically the number of stress, MPI, stress echo performed in low risk patients with the following exclusions and the exclusions go on and on and on.

Patients qualify for this numerator if asymptomatic and low risk and not any of the following and any of the following are, they've had a stress echo, they've had MPI, and it goes on. I can't understand who
could get into the numerator after all these exclusions.

It seems like we're not getting at the issue of inappropriate use of stress MPI. We might be identifying one or two patients who had inappropriate stress MPIs. I have real problems with this measure.

DR. GRIFFEY: Can I comment on that? Richard Griffey. If you basically did, as someone suggested, and said prior history of CAD or prior testing, then you would cover all of those without enumerating them and that would make it very interpretable. It's kind of like that other -- when we listed all the criteria out of the Geneva score or whatever.

DR. SNOW: Snow here. I don't agree with that. I think that one of the problems that this -- as structured, that this creates is, again, you can tick off the list but as you describe you would still be in the situation that if you had done calcium scoring, which is probably not justified, you
then drop off the numerator.

    I haven't faced this before but as this has come up and I've been thinking about it, I think this is probably almost categorically not a good idea to have procedures or tests of excludable items because that just encourages the guy to do something he shouldn't be doing. That should be a criteria. If you take the more general no-known cardiac disease, well, that's a little bit different.

    CO-CHAIR PETERSON: But also vague.

    DR. SNOW: It's still vague but one of the things about vague assignments is that psychologists -- why do we want psychologists in here? The psychologists who study this say that if you ask people to give you information about something they will say, I can't do it. It's too vague. Of course you would say, do it. Make the choice. One, two, three, four, five they give a Gaussian
distribution.

CO-CHAIR PETERSON: It's pretty hard for a measure, though.

DR. SNOW: Yeah but, I mean, vague information is not necessarily bad information is the point.

DR. GRIFFEY: Richard Griffey.

This is done all the time though in measures where you'll say no cardiac disease asterisk and at the bottom it lists the exclusions without making it hard to interpret but making it very usable. I don't think it's that big of a deal personally.

DR. RUCKER: I think one of the challenges here is if you look at sort of the cheapness of snip chips and some of these other technologies, we are extraordinarily close to having multivariate very high fidelity predictions of pulmonary and fairly genetically determined diseases in large part given some of the expression things.

This sort of has the smell of
something that's going to go away pretty
quickly and seem obsolete to me. As you look
at these things they are moving very, very
rapidly and then you are actually going to
have I think -- right now they are sort of
comparable to Framingham, you know, but I
think they are going to exceed that.

If you're going to use a
Framingham that's a big data collection
workflow challenge for people to start looking
up on charts and calculating things. I
thought I would just throw that out.

MR. BACKUS: This is Mike Backus.
We struggled with this issue in our pre-auth
process. We kind of narrowed it down to five
or six factors that essentially -- as compared
to the 12 or 13. It amounted to cholesterol
and diabetes, smoking, five or six to try and
streamline that down.

Then essentially what we did is we
assigned points to that so as you looked at a
patient if they were high risk depending upon
what the ACC guideline was they essentially automatically qualified. I don't like the measure because of feasibility.

From a clinical perspective I don't see that 30-person thing as a problem because what it says is just go back and look at the last 30. Any institution is going to have done 30 stress procedures be it MPI or stress echo. I'm not sure, however, this is going to go away when you look at the installed base of equipment and investment that everybody has, I think it will be a long time until this goes away, CMS CAMEN performance notwithstanding.

CO-CHAIR PETERSON: Try to keep new comments going. I think people are generally settling in on opinions here.

DR. CANTRILL: Steve Cantrill. 2(a)9 and 2(a)10 I'm concerned that basically poor documentation is being rewarded by excluding the patient. I think that, to me, is just as bad.
DR. SPENCER: Four quick comments.

One is the stress MRI should be included. The second one is I like the measure, again, because this is a real problem and it's an important problem, although, again, we haven't solved how to measure it. As written not great but, again, clearly agreeing that this is a problem.

The exclusion criteria, let me just defend those for a second. I think that big long list is meant to say two things. One is the first half of what you said, their known history of coronary heart disease, so we can just get rid of that.

The second half is there because of the initial assessment. I'm not sure why this had to focus on the initial assessment of an asymptomatic patient. I mean, the second time when an asymptomatic patient is being assessed it's even worse than the first time.

DR. SPENCER: I'd love to hear the logic of why you would want to focus on
initial assessment, but that's what the second
half of those tests are for because their
exclusions have already been assessed.

Lastly, again, I just don't know where
the data comes from. If this is coming from -
- I can't believe that we are going to review
data charts, clinic charts and requisitions.
Maybe I'm just naive but five things that say,
fourty-two year old guy doing great. EKG
normal. Plan stress. And then the req.
You've got to check a box on the req and the
req has boxes on it. You can really submit a
req to your stress lab that they'll take that
says, none, asymptomatic.

DR. BURSTIN: It happens all the
time.

DR. SPENCER: Not with imaging.
They get EKGs. They don't get echos. I mean,
we don't do those. That's why I like the lab
measure.

MR. BACKUS: This is Mike Backus.
I can tell you from our data, and I'm sure NIA
would corroborate it, we have a huge focus in cardiac right now as lots of people with the ACC know. In stress echo, in MPI we routinely see between four and five percent of the requests that come in the door that after discussion the physician either withdraws or not. In stress echo, frankly, we see it higher. When we look at the demographic data of the patients, what we tend to find is that MPI is ordered more on older men, stress echo is ordered more on younger women. Relative to the risk score or pre-test probability, MPI tends to be ordered on essentially sicker men and stress echo is essentially ordered on healthier women. In the stress echo realm we see higher withdrawal rates and stuff than that so I would tell you that the ACC in their pilots came up with numbers 15 percent inappropriate, 15 percent questionable, 70 percent all good. Those are more stunning numbers to me than we come up with when we do pre-auth. We see it. We see it all the time.
DR. SPENCER: Is this fixable?

DR. GRIFFEY: If you just said --

DR. SPENCER: Do we think it's a problem?

DR. GRIFFEY: If you just said no CAD you wouldn't get at everybody but you would get the tip of the iceberg. Right?

DR. SPENCER: Low risk and no known CAD.

DR. GRIFFEY: And if you fix the denominator, meaning get rid of the exclusions for no information, would we have something that we could approve?

DR. SPENCER: The only other thing would be the assessment of low risk. How do you determine low risk? It sounds like --

DR. SPENCER: If we demand Framingham the feasibility goes to --

DR. GRIFFEY: The question is is something better than nothing.

DR. STILLMAN: Although, as Eric pointed out, you tend to overestimate risk
rather than underestimate it so it might not
be such a bad thing.

DR. GRIFFEY: If someone is
asymptomatic I guess they could still be
determined to be moderate risk.

DR. CANTRILL: If the doc says
they're low risk, yeah, maybe they're no risk.

DR. SNOW: This is Snow. The
tendency is to try to come up with something
that will get to the test. You've decided you
want the test for whatever reason. Maybe the
patient said he wanted the test so, okay. How
can I get this for you, Joe? Well, give him
medium risk.

CO-CHAIR PETERSON: Joe, in the
interest of time you've heard the major
complaints and comments about suggested
changes. Do you want to make some general
comments and then maybe we'll have you respond
to a few specifics.

DR. ALLEN: Sure. In general,
people are saying it's an issue and it was the
most frequent inappropriate indication that we found in our pilot so that's one of the reasons why we are putting a measure forward in this area. Not to focus on it, not look at a frequent inappropriate, a large percentage of the inappropriate.

The exclusion of not recorded, it's just a matter of fact that when you go to do these measures some data will be missing. The original PERC measure was found that if you under-reported you would come out on that measure poorly. That had to come out and we could see that but you'll still be adding those patients back in and increasing your denominator making your inappropriate rate look better.

We would still encourage you to keep that exclusion in because at least your denominator becomes smaller. Even if you don't have the data for whatever reason, it reports the true inappropriate. If you don't have the data you can't report it or look at
it so it would just automatically fall in.

There is a denominator to make your work better.

Then the reason why we focused on initial was because there are concerns that once you're past and you say you do have a health component that is really high, you are no longer in the initial risk assessment period so there is a difference there and how do you handle patients where you already have some information? They are no longer truly patients that can be assessed by Framingham because you have additional data now.

CO-CHAIR GAZELLE: If I could comment on the denominator exclusion. I think what our feeling is they should not be excluded from the denominator but they would also then end up in the numerator as being inappropriate.

DR. ALLEN: You wouldn't have enough information.

CO-CHAIR GAZELLE: Right. So they
are inappropriate. If you don't have the
information, they are inappropriate.

DR. ALLEN: You don't know that
they might have been sent for some other
reasons that we are not tracking because there
are more reasons such as PCI asymptomatic
patients and the peri-op patients. They could
be symptomatic but you just don't have the
information.

CO-CHAIR GAZELLE: We would agree
that part of the hope here is that we are
documenting the reason for appropriate
imaging. If it's not documented, then it's
inappropriate. It's inappropriate imaging.
You got to have some reason for it.

CO-CHAIR PETERSON: And since the
public reporting of this there will be a
strong indication for sites to appropriately
document why they are doing what they're doing
and don't have the reason for any of it.

DR. SPENCER: That's the way it's
all done. I can talk to a patient for an hour
about smoking cessation. If I didn't write it
down, you know, I get dinged on it.

DR. ALLEN: So it would get added
back into all of the numerators for each of
the three measures we're discussing then? If
you don't have enough information, it depends
on how much information I guess you have on
which should go in but, I mean, if you have
very little information, their symptom status
or things like that, it just goes into all
three, or you don't have enough history on
PCI. That's the challenge that we have. How
do you add them back in and make them
inappropriate because you may end up double
counting some in the measure.

CO-CHAIR GAZELLE: I don't see how
they are double counted. So if somebody ends
up in the denominator for having had one of
the exams and they end up in the numerator for
having had one where there are not documented
indications essentially for it being
appropriate so it's not double counted.
DR. ALLEN: It would be counted in this measure and the PCI measure and the peri-op measure.

CO-CHAIR GAZELLE: We're just considering one measure right now. We're not voting on them as a block.

DR. ALLEN: Correct.

MR. BACKUS: Right, but he's saying the same problem exists in PCI.

CO-CHAIR GAZELLE: But we haven't approved it.

MR. BACKUS: PCI, I thought we did.

DR. BURSTIN: We already did PCI.

MR. BACKUS: We already did PCI.

CO-CHAIR GAZELLE: It goes for recommendation, right? We shouldn't reuse the fact that it might be also included in the other one to support this measure.

DR. GRIFFEY: No, he's saying the other way.

DR. BURSTIN: I think he's saying the other direction.
MR. BACKUS: And I think he's probably right but we're not talking about that.

DR. BURSTIN: Right. We should probably revisit that exclusion in that measure.

CO-CHAIR PETERSON: Another proposal I think you mentioned would there be a problem in expanding this to MRI?

DR. ALLEN: To MRI?

CO-CHAIR PETERSON: Yes.

DR. ALLEN: No.

CO-CHAIR PETERSON: Okay.

Was there anything else on yours, Kirk?

DR. SPENCER: I don't know that we fixed the two that -- no, nothing new.

CO-CHAIR PETERSON: Were there other things that we could -- well, I mean, I guess the point is is there anything else you want him to address or are we going to vote on what we have? Is there anyway else to improve
the measure, at least on your belief, or want him to address answers to why we don't have something better?

DR. BURSTIN: I think if we actually think there is something we wanted to improve then, again, I think the resubmitted measure should come back in and we should actually vote on that because I think it's probably substantially different.

DR. SPENCER: It might be helpful to have committee discussion about the initial assessment again. I mean, you give an example of calcium scores. I think that is a very unique situation. I mean, prior testing, that gets covered because that's prior evidence of CAD so that is not a good example. Right?

DR. ALLEN: First, I think you could have a clinical stress test, which often happens in exercise, the standard without imaging so now you have a patient who you have an initial global risk assessment but do you have a non-determinant first time test?
DR. SPENCER: That's a code for a test. What if your prior testing was normal? I'm just trying to get rid of normal testing on top of a normal test.

DR. ALLEN: We do have appropriate use. We didn't know the measure around the repeat testing because we didn't find in our pilot that was coming up as frequently as the things that we put forward. An additional measure for repeat testing could be a target.

DR. SPENCER: So your prior test was equivocal so you didn't have CAD definitely. If it was normal, I just don't know why a prior normal echo makes it okay to get an inappropriate stress nuclear.

DR. ALLEN: It wouldn't.

DR. SPENCER: It would here.

DR. ALLEN: -- criteria as being inappropriate. We just aren't measuring it with this particular measure.

DR. SPENCER: Okay. It wouldn't catch everybody but it would catch some so
that doesn't fake the number, we just lose
catching it. Okay.

CO-CHAIR GAZELLE: Everyone agrees
that overuse of stress imaging is a big deal.
I'm just not convinced this is going to catch
much of it so I don't think it's of value.

CO-CHAIR PETERSON: Okay. Any
other comments internally? I guess the
options now are two. One of them would be to
say we need to have substantial changes to the
measure that we currently have and have a
revote of this internally or we say, no, we
have enough information here. Substantial
changes would not change how we would count
the vote today and we move it forward to a
vote.

DR. GRIFFEY: Specifically the
changes would include the risk assessment, the
exclusion criteria, and this issue with the
denominator data.

DR. BURSTIN: In addition to MRI.

CO-CHAIR PETERSON: In addition to
MRI. So let's have a vote. Just to be clear, the vote is substantial changes versus no, we want to vote on this today. If we want to vote on this today, then we will accept only minor changes. I guess the MRI could theoretically be put in there. Everything else could be more substantial and we would vote on it up or down based on that change. Does that work?

DR. GRIFFEY: With some of the other measures we've had a similar change.

CO-CHAIR PETERSON: Right, but would there be interest to receive that?

DR. BURSTIN: Do you want to essentially have this measure attempt to come back with substantial changes which was outlined and then review it at a later date or not? Is it fixable?

CO-CHAIR PETERSON: Okay. How many would vote that it's reasonable to reconsider the measure after substantial changes?
DR. BURSTIN: Okay. We're done.

CO-CHAIR PETERSON: Great. So you
are going to get the substantial changes
requested.

DR. SNOW: With no guarantee.

DR. BURSTIN: All of it is just
recommended conditions. The conditions come
back to you. You review it again and you
make --

CO-CHAIR PETERSON: You would have
to resubmit and say, we don't care to do that.
Okay.

Next. Two more. Okay. We are
at --

DR. SPENCER: We could clear out
the ACC.

CO-CHAIR PETERSON: Yes, clear out
the ACC and let's get Joe done.

DR. FIESINGER: This is Troy
Fiesinger. I'm got the pre-op ACC and then
Don is doing the pre-op CMS.

PARTICIPANT: Could you give us a
number, please?

DR. FIESINGER: It is 14-10.

That's what we're on now. They are very similar I'll warn you ahead of time. For this measure it is essentially looking at cardiac stress imaging in pre-op evaluation of low risk surgery patient. For example, cataract surgery, there is a long list of others, endoscopy, so on and so forth. Something I do every week doing these kind of clearances in the office. So the numerator is number of stress echo cardiograms, SPECT MPIs in low risk surgery patients as part of preoperative evaluation. That's your numerator. Your denominator is the number of SPECT MPIs stress echo cardiograms performed overall.

The second measure has a different denominator that looks at the same issue in a different way. So to go through it item by item, 1(a). High Impact. I gave it an M as in Martin and the question I have is overall volume of these procedures as a percentage of
the total number of SPECTs and echos because I very rarely ever send any of these low risk patients on for additional imaging. If they do, they've got symptoms or have very low MET scores or things like that. That's what I was asking you earlier for overall prevalence data.

CO-CHAIR PETERSON: Joe, I don't know if you want to give this as well.

DR. ALLEN: Sure. The quick answer, it is variable by institution. Some institutions have a high volume of this and have an issue with this. Others don't based on the referral pattern so there would be variability both on referral patterns but also in the amount of time that they've been productive. In some cases there will be a very low volume of peri-op cases. In other cases it was the number 1 and number 2 issue.

DR. FIESINGER: In terms of overall inappropriate SPECTS and echos how does it compare to the asymptomatic patients
getting and opposed CABG or PCI patients getting them. How big a slice of the pie is this piece?

DR. ALLEN: Again, it depends on the institution. For some it was the biggest piece of the pie. Others it was the smallest piece of the pie.

DR. FIESINGER: From the nuclear --

DR. ALLEN: In the top four.

DR. FIESINGER: Of a percentage of inappropriate studies the asymptomatic low risk CHD was 45 percent, the post-PCI was 24 percent, the pre-op was 3.7 percent.

DR. ALLEN: In our pilot we had all outpatient cardiology practices so we didn't tend to deal as much with hospital-based referrals from anesthesiologists so it was very low in our pilot. Other studies have it much higher when they had a closer association of the lab with anaesthesiology especially in outpatient and hospital time.
DR. FIESINGER: My past experience is low risk surgery patients generally we're doing in the outpatient setting. What I'm doing is based on the hospital is high risk surgery or intermediate risk surgery and a high risk patient. This seemed to stratify pretty starkly but thanks for the information. 

For 1(b), Opportunity for Improvement, I gave this an M also and that ties into the low overall percentage of this inappropriate test in view of all the tests.

In terms of 1(c), Relationship to Outcomes, I agree the evidence says these shouldn't be done and the standard should be zero so I did give that a C.

When you come to the end it was Threshold Criteria of Importance Met. I hesitate to say yes based on the overall prevalence. If we are going to drive this to zero is that going to impact the overall problem as much as driving the other numbers to zero in the asymptomatic patient as opposed
to a CABG patient. I personally think it should never be done. I agree completely but I'm not sure this is where efforts should be focused.

Do you want me to stop there or keep going?

CO-CHAIR GAZELLE: Vote on it.

CO-CHAIR PETERSON: I think that we need discussion.

DR. FIESINGER: I welcome other people's opinions. I'm really on the fence. Personally I think one thing we're looking societally and then towards the healthcare system. I hesitate.

DR. RUCKER: One additional data point on the 10 things that Troy and I talked about was that in the numbers on the CMS on the Lewin Group data, I think the average -- they were saying there is a big issue on what exactly the mix of low risk is because there are two very different definitions of low risk in these two measures. The incidence of the
study in low risk was .005 on average so that
would be one in 200 Medicare patients as I
understand the math in the Lewin analysis are
getting that, the top 1 percent. The heaviest
users I think are roughly one in 40 so you are
plowing through a lot of old folks who
presumably, if any of them have any kind of
sort of little symptoms, especially when you
consider that over the 30 days what Troy
looked at was 60 in 30 days. If you are 65 to
85 you have -- roughly if you live 20 years on
average into Medicare you have 240 month-long
cohorts of 30. If you have one stress just in
general, you are basically at .005 as the rate
even skipping any surgery or relationship to
surgery. Some of these numbers look, at least
by this, look to be very low. I think neither
of them had risk stratification of the patient
as opposed to risk stratification of the
procedure. I would just throw that out.

CO-CHAIR PETERSON: It's
interesting. I was just looking up some of
the references that were given, the Mayo study
done by Dr. Gibbons. It appears, unless I'm
misreading this, that other than asymptomatic
testing in an asymptomatic population, low
risk population, that pre-op testing was the
second most common and very close in its order
of magnitude. Of the inappropriate testing
this was the big one.

DR. RUCKER: So that's the Mayo.

CO-CHAIR PETERSON: That's the
Mayo clinic. I can find a few more here.

DR. BURSTIN: So you're saying
it's a big problem potentially?

CO-CHAIR PETERSON: I'm thinking
according to this it is. I think actually
pre-op testing in general is one of the big
abuse areas that could be cut back pretty
easily.

DR. RUCKER: So then we should go
on.

CO-CHAIR PETERSON: Mostly the
issues of people not being eligible by the
pre-op guidelines.

DR. BURSTIN: This is also probably a pretty significant bias in generalists versus specialists so I think if you're at a cardiology practice versus a primary care practice, the access and the likely utilization of these tests is very different.

CO-CHAIR PETERSON: Everybody in-house gets tested because we feel they sent them to us for pre-op evaluation.

DR. GIBBONS: This is a very serious problem because the number should be zero. I'm torn a bit because I see the problems you are talking about here but having a measure as a starting point for management, we don't have it. We all know it's out there. An interesting point that it seems to vary a great deal from institution to types of institutions. I think that is important. It probably varies across other elements or domains of this section, too.
MS. ZERZAN: This is Judy. I think one question would be this measure versus the CMS measure. I think something has to be done in this area, I totally agree. One, both, who knows is the question.

DR. FIESINGER: We have that question, too. We'll bring that up at the end. I'm going to continue this.

So we go to 2(a) Measure Specifications. We already talked about the 60-day time period, ACC folks thought that replicated 12 months. Initially I was happy with the numerator being just factoring in PCI but I do need to add stress MRI. That is a question.

DR. SPENCER: I can't imagine anybody doing a CTA for pre-op assessment but, okay.

DR. FIESINGER: For the denominator the same issue. They have the same exclusion, insufficient data. I would add those patients back in considering
previous discussion.

In terms of testing and analysis I gave it a N because it said no direct reliability testing done. This is 2(b). I don't think that is a make or break but it sounds like more testing should be done with this measure.

2(c) I gave it a C. To me it's clear in the evidence that you shouldn't do this. Summary of evidence supporting the exclusions is 2(d). I think we need to kick out the inadequate data and exclude them so I gave it an M based on that.

In terms of 2(f). I gave it a P with a little bit of a question. The answer is probably with a wide variability maybe there are some high outliers we can chase to get corrective behavior so I'm okay with that.

One thing I threw in there, too, is there was not risk stratification of patients. Risk stratification was solely surgery type which is a standard question.
When I do this I'm doing it at least by risk index, I'm asking other questions from the guidelines. What are your thoughts about patient risk being put into this or why did you leave outpatient risk?

DR. SPENCER: I mean, for minimal risk surgery patient risk doesn't enter in. You just can't have unstable angina or unstable arrhythmias being put in with heart failure. Basically you have an acute MI and MBT or heart failure you don't have to risk assessment in the new guideline. You don't even get into the mets if it's low-risk surgery. That's what one of the changes was. Low risk is low risk.

DR. RUCKER: None of that is actually asked in either of these. So, you know, active angina or something is not --

DR. SPENCER: Right. The patients --

DR. RUCKER: -- or assurance of --

DR. SPENCER: Whatever the table
of guidelines is.

DR. RUCKER: Which wouldn't be a problem except for the fact that you have these very low Ns, one in 200. You can certainly manage the one in 200 in elderly patients that are having some kind of active cardiac.

DR. SPENCER: All the things on that list is grossly inappropriate to send them to a stress test, too. So that's why it kind of doesn't matter for this measure. People with pulmonary heart failure, MBP, probably shouldn't go for a stress test. Probably shouldn't go to the stress lab either.

DR. FIESINGER: Okay.

DR. SPENCER: It's a new measure, you know.

DR. FIESINGER: We'll go to number 3, Usability. I gave it a P. That's probably based on my question about the prevalence. For 3(a), rather, I gave that a P.
3(b), harmonization, I gave that an M. To me the issue is the CMS measure which is extremely similar and we need to address that. 3(c), the Singular Additive Value, I gave it a P. It's the same issue. There is a companion measure that is very similar.

Overall I would P. We have to address the issue of the second measure and how to handle this. In terms of Feasibility, 4(a), I gave it a P. 4(b) I gave it a P but it's going to require a paper data capture tool, at least by what I saw here. That could be a bit of an issue if that's how you propose to do it. So I would wonder if it can be done solely by administrative data. Overall Feasibility I give it a P.

Overall recommend for endorsement, I think I would put that to a vote. It's the prevalence question that just sticks. In principle I agree with driving this to zero completely.
CO-CHAIR GAZELLE: Okay. Comments from the group?

Helen, maybe you can address this.

I still have problems with this 60-day sampling period as opposed to a whole year. I mean, it seems that most of these measures are burdensome and, yet, most of them we require a year's worth of data. Is there a precedent in the NQF for other measures that have sampling?

DR. BURSTIN: It's whatever sampling frame is appropriate as long as it's justified.

CO-CHAIR GAZELLE: Are there other measures that are readily chosen?

DR. RUCKER: Was it 60 days or 60 days within surgery? I thought it was 60 days within surgery as opposed to 30 days within surgery.

PARTICIPANT: For the ACC it's a 60-day sampling period.

DR. RUCKER: Okay. I missed
surgery.

CO-CHAIR GAZELLE: It's 60-day sampling period. Select a starting month; January, March, May, etc. Begin a 60-day sampling period and you must have at least 30 stress echoes or SPECT orders. It's fairly low.

DR. BURSTIN: It's a reasonable number to measure that.

CO-CHAIR GAZELLE: But it's a fairly low threshold to get 30.

DR. GIBBONS: But these pre-ops are the rates you're talking about?

CO-CHAIR GAZELLE: Thirty numerator events.

DR. BURSTIN: That's a lot. It's huge.

DR. GIBBONS: Unless you have some real outliers for ordering it on every cataract or every colonoscopy you're not going to get that.

CO-CHAIR GAZELLE: Other comments?
DR. SPENCER: This is Kirk Spencer. I think we let them off the hook too easy. I would like to hear the justification on the pre-op. Apparently if you put any other indication for the stress you don't get in the numerator or you don't get -- yeah, you don't get counted at all.

You have pre-op and you check the box that says remote history of CAD or abnormal EKG this measure doesn't count you. They should count. I don't care what other symptoms you have. If you're going for a low-risk surgery and you are being pre-oped there is nothing else you can check on the box that makes it appropriate I don't think. Is there? ACC?

CO-CHAIR GAZELLE: Joe, can you respond?

DR. ALLEN: Sure. There is no other reason as recorded for the imaging because if you have a high pre-test probability patient who would have otherwise
qualified now, you know, this comes to a feasibility issue. Can you get the additional data? We did in this case leave it more vague. There are other reasons that a patient could show up that would be a justification for the test.

DR. SPENCER: There are some other reasons to have a stress test coincidentally 30 days before your surgery but I thought I read that literally is what the requisition said. The requisition says pre-op and abnormal EKG. Right? You're saying they let the people off the hook who are trying to -- then don't check the pre-op box. Just check the box that says abnormal EKG. That's my point. You're checking pre-op on the box, then you need to make sure it's valid. Maybe that's an education issue. People see their score is high and they say, I better just check chest pain and not check the pre-op box. Right? That's what I'm getting at. They say pre-op. Shouldn't we ding them?
DR. ALLEN: The primary reason is just not record peri-op, that's it.

DR. SPENCER: Well I second that, I don't care.

CO-CHAIR PETERSON: Just as a clarification of the sixty days, it's not the numerator. It's the denominator and so given the rate you could get to 30 stress SPECTs just because of sampling because the rate is so low.

DR. CANTRILL: Isn't that true of the previous one, too?

PARTICIPANT: I had a problem with the previous one, too, but we didn't like the previous one for other reasons.

CO-CHAIR PETERSON: So, to clarify, on both of these we would want to have notes to sample the year. Were there other comments on this one?

Joe, you haven't spoken yet since some of the issues that you heard were on the table. Do you have more things to say?
DR. ALLEN: I've heard a lot about the volume and whether or not this is an issue. I've spoken to that and there has been a number of comments from the table. Most of the other comments were -- stress MRI, CT.

DR. RUCKER: Joe, the one question -- Don Rucker -- was your definition of low risk in this measure, and not in the CMS measure, is 1 percent MRI or mortality rate? To me that seems like a pretty high -- my understand is that well-done CABGs have a mortality rate now under 2 percent so I'm thinking 1.9 percent in some procedure, let's say, morbidity, mortality rate is actually fairly high risk procedure.

That part of the definition seemed a little -- it seemed like there would be some better definition for that or some clarification. In part this is motivated by knowing that on the CMS list, there is a very heterogeneous list of procedures with orders of magnitude difference and risk on that list.
DR. ALLEN: The definition is taken from the guideline. That is what we chose, based on the guideline literature, to support the guideline. There were other data to support the different kind of points. We could go into it further, but that's why we chose --

DR. RUCKER: I mean, general anesthesia was very proud of themselves when they went from one in 10,000 mortality to one in 30,000 when they instituted pulse ox just to put one in a 100 or, if you will, whatever that is, 110,000 into perspective in terms of risk.

CO-CHAIR PETERSON: I think it's very unclear.

DR. FIESINGER: Quick question I didn't ask earlier. Is sinus testing an issue or is this all tests done anywhere meaning the patient imaging department versus a free-standing facility?

CO-CHAIR PETERSON: This is
anywhere.

DR. FIESINGER: The CMS data we looked at shows decline in testing use in hospitals outpatients.

CO-CHAIR PETERSON: So the options here, and we're at a juncture point unless there are further comments, we can either break now and review the CMS measures which is the same topic but a different dataset in a slightly different variable and then vote on both. Or we could have this one revised and brought back to the group. I don't think we are ready to vote on this. What would your pleasure be?

MS. ZERZAN: CMS.


DR. RUCKER: Let me say very quickly since I think we discussed everything. Impact, I think we discussed that at length. I would say that is sort of a P. SPECT in and of itself is clearly an issue. The pre-op I defer to the numbers that were cited.
1(b) Opportunity for Improvement,

I would put that as a partial too. Again, my concerns here were in the Lewin Group analysis the rate that they cited, .005, you know, as a ratio of pre-op past here to low-risk surgery that one in 200 patients. I just wonder when you have something that is one in 200 so 199 people aren't doing it and one person is doing it I just wonder what -- I think there may be some reasons that are quite rational as opposed to just --

DR. BURSTIN: Was the analysis again limited to only hospital outpatient departments on this one? So it would significantly expand the numbers if you went beyond hospital outpatient departments I assume. Right?

DR. RUCKER: It's just a ratio.

DR. BURSTIN: It's still only including hospital outpatient.

DR. RUCKER: Okay. All right. We talked about that, 1(c), Outcome or Evidence
Support Measure. I was unclear about that.

I guess I have to give it an N. As far as I can tell from the references here there wasn't any actual research cited. These were, as far as I can tell, all guidelines so this was a series of various guidelines.

DR. BURSTIN: Most of our evidence is based on guidelines. That's actually quite appropriate.

DR. RUCKER: Okay. All right. So they were based on guidelines.

DR. BURSTIN: High quality guidelines, our stock in trade.

DR. RUCKER: Okay. If that is your stock in trade you've got more stock.

On the threshold importance, I guess if we take the numbers that Kirk gave it was probably met. On 2(a), Measure Specification, this was a pretty straightforward measure. It's declined. The numerator, number of stress echoes SPECT MPI and stress MRI studies performed at the
hospital outpatient facility in the 30 days preceding low-risk non-cardiac surgery.
That's pretty crisp. The denominator, is also relatively crisp. The number of low-risk non-cardiac surgeries performed at the hospital outpatient facility. Just to point out, there is no risk stratification of the patients whatsoever including even the active MI after heart failure. The way that the denominator is done here --

DR. BURSTIN: They are all low risk, so we have already --

DR. RUCKER: No, no, no. The surgery is low risk. The patient risk is totally unspecified. They could be one second from, you know, demise.

Did the denominator detail? 2(a)8 is several pages of procedures most of which are, you know, upper and lower endoscopies, some ENT procedures, some opto procedures. There are some interesting things in here. I don't know if vascular endoscopy is.
Some of these procedures seem pretty high risk. For example, there is laposcopic gastric bypass with Roux-en-Y. If you are getting a gastric bypass procedure with an Roux-en-Y, that strikes me as different than an endoscopy for a nasal cauterization which is also in the same risk.

DR. SPENCER: You realize you're 450 pounds if you're having that done.

DR. RUCKER: Right. Exactly.

This is an assigned mix by what is given here different than just the 1 percent or 1 to 5 percent ratio.

Some other things; pyelotomy. One of the very lowest rhinoscopy, epistaxis control which I think is a swizzle stick endoscopy which I generally don't do under general. Those are the procedures. Risk adjustment we talked about. No risk adjustment.

2(b), Reliability testing. I think here there is the same text that shows
up in a lot of these which is minimum case counts were developed to ensure 98 percent confidence level. Case counts requirements range between 31 and 67.

Again, the echo raised comments, I don't understand the statistical validity of requiring 31 to 67 patients when you have something that averages one in 200 and the worse case, the absolute worse case, I think, was cited as the top worse 1 percent was around one in 40 or one in 35. Those numbers seem off to me just from a raw power conclusion.

2(c), Validity, I think, you know, the claims as claims are -- I would say that's a C. Exclusions there are essentially none so I made that NA. Risk adjustment we talked about is the end. 2(f). Identification of Difference in Performance, again, is the same issue with the number.

2g, Comparability, NA. Disparities, NA. Overall scientific
acceptability because of the issues and the
risk I would say that is an M. On the
usability I thought other than the risk
stratification I thought it was a P so that
would work.

Harmonization, obviously Troy
discussed. Distinct added value, again,
that's sort of the same as Harmonization.
Feasibility, I would say those are all Cs to
me other than 4(d) because I think it's pretty
feasible to do off the claims data for
Medicare. I think that is a very
straightforward thing. That risk procedure I
would make it a P on 4(d). The other is 4(d),
also a C. I guess that's all I have to say.

CO-CHAIR PETERSON: Okay. We have
an issue. We are going to lose two more of us
now. I've got a 4:40 flight.

DR. SMITH-BINDMAN: Have you
thought about importance?

CO-CHAIR PETERSON: I guess the
question is we could vote about importance.
How many think that this measure meets importance?

PARTICIPANT: Just this one or the two?

CO-CHAIR PETERSON: Just the one. I think it's going to make importance unfortunately. I don't think it's going to be dinged on that. It's going to need to go to question and I think we need to have another call to finish this and the other measure. It would be unfair to CMS not to and we could invite them and ACC both on that call.

(Whereupon, at 3:15 p.m. the meeting was adjourned.)
| Page 307 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Page 308 | Neal R. Gross & Co., Inc. | 202-234-4433 |
Neal R. Gross & Co., Inc.
202-234-4433
errors 16:3 123:15
especially 43:21

establish 141:10
establishment 249:21
estimate 65:17
249:5,10
estimates 9:16
European 10:12
11:10 18:7
evaluate 217:8
242:5
evaluated 61:1
117:3
evaluating 117:10
148:4
evaluation 8:4
10:17 39:14
175:22 179:17
276:6,14 283:11
event 23:14 34:17
219:22
events 290:15
eventually 83:3
109:1
everybody's 76:12
evidence 10:9
17:15,16 30:20
50:1 53:8,18 54:5
54:16 83:19 84:1
84:3,9,14,17,22
87:3,13 88:2,10
93:14,17,19
101:19 102:11,15
103:1 107:13,14
118:4,17 119:14
146:22 184:20,20
185:5 244:5,7
246:15 271:15
279:13 285:9,10
297:22 298:7

evidence-based 49:14

evident 148:10
evolving 34:21

exact 33:15 56:19
138:4

exactly 6:22 26:1
38:5 72:10 125:4
178:15 195:12
237:15,16 280:20
300:10

exam 41:10 153:15
242:19

examination 85:8
104:22

examinations 119:3

example 74:20
119:2 129:13
143:8,9 158:4
169:17 240:22
247:22 271:12,16
276:7 300:2

exams 62:1 154:17
268:19

exceed 258:7

exceeded 147:4

exceeds 147:11

excellent 92:5

exception 83:12
188:15

exceptly 52:7

exceptions 176:17
202:12

excess 171:16

excessive 59:10

excludable 256:6

exclude 36:6,15
42:12 186:5
88:21 189:3
193:10 285:12

excluded 13:18
36:5,10 54:4
88:15 111:17
123:20 151:2
158:7 161:10
162:19 164:18
266:17

excludes 149:2,8,10
149:13 158:12,14
160:9

excluding 55:18
97:9 151:14 159:9
246:6 259:21

exclusion 36:16
106:8 118:10,19
119:10 120:4
126:3,8 131:21,22
132:11,22 143:15
151:3,6,8 153:21
154:13 157:21
178:5 179:19
209:2 245:4,12,21
246:16 247:19
260:9 265:7,18
266:15 270:5
273:19 284:21

exclusions 13:12
14:16,18 16:2
55:16 94:16 100:5
101:16 104:2,7
111:16 120:2,3
121:13 123:13
125:17 131:17
134:17,19 139:16
139:17 159:7,18
163:7,14 168:20
168:22 170:22

172:22 176:9,20
179:22 180:7
186:18 187:10
210:6 211:1 231:8
236:6 237:10
244:13 247:16
250:7 254:16,17
255:2 257:10
261:3 263:11
285:11 301:16

Excuse 156:7
205:16

exercise 271:19

exertion 105:12

exhaustive 101:15

exist 36:1 76:19,20
186:13 209:2

existing 57:4 82:17

exists 269:9

expanded 62:6 233:6
297:15

expanded 97:16
232:22

expanding 177:1
215:18 224:17
225:14 270:9

expansion 224:15

expect 136:9

expectation 235:8

expectations 237:5

expecting 19:2
217:10,12 218:8,9
218:11 221:18

experience 53:15
60:19 67:15 130:5
193:13 195:16
279:1

experienced 58:1

experpal 132:21

expert 88:1 94:18
135:16,20

experts 93:19

explain 210:3
250:20

explicit 246:3

explicitly 36:6
45:12 103:7

exposure 65:12

117:22

express 38:15
127:3

expression 257:21

tend 215:8

tension 80:5
130:9
tensive 94:5
171:3 174:11
251:22

tent 55:21 61:13
177:3

extra 51:9

extract 34:11

extracted 123:8

extracting 15:22

extraordinarily 257:17

extreme 216:5

extremely 288:3
eye 45:15 46:16
eyeballs 66:11

F

F 112:16
FAAFP 1:16
FAAOA 1:20

face 57:15 108:20
179:2,7,14 210:22
211:19

faced 256:2

FACEP 1:15

facilitates 121:6

facilities 134:6
136:5,19 138:2
139:12 156:17
173:5 177:2 199:3
212:3 247:14

facility 128:8
133:11 136:7,17
139:1,6 150:14
153:1,6,11 154:16
167:13 295:21
299:1,6

facility-level 123:11
Page 319
202:19 204:19
249:16 257:3,11
harder 96:1
harm 151:17
203:12
harmonization
15:14 72:10 96:4
123:3 176:3 187:3
247:5 288:1 302:6
302:8
harmonize 177:5
harmonized 57:7,7
57:10,10,12,14
70:15 174:20
175:2 233:17
harms 179:9
harsh 248:14
hate 248:14,21
hazard 249:15
head 21:18 30:11
31:12 48:15 49:13
51:3 53:8,22 58:4
58:9 59:1 60:20
61:2,5,7,9 63:11
64:11,14,16,17,19
81:17 82:2,7 99:2
104:17 110:7
116:19,20 117:4,6
118:9 119:1,20
121:14 122:9
125:22 130:6,10
130:12 131:2
135:5,8 145:12
219:9
headache 48:16
81:18 82:3 83:1
83:10 85:7,10
87:15 90:5,9,15
90:22 92:8 97:18
103:4 104:20,21
105:8 106:12
111:10,20 117:3
117:13 118:6
119:3 129:20
131:13 135:1,19
headaches 82:8
93:1 97:8,9,10,12

97:13,22 105:7,10
105:11,12,15,19
113:17 119:4,5
129:19 130:14
heading 113:17
headline 180:21
head-only 61:4
Head/Spine 3:4
health 96:2 103:5
266:7
healthcare 155:16
164:8 243:20
280:13
healthier 262:15
hear 86:11 171:13
260:21 291:3
heard 107:3 185:6
200:21 237:3
264:16 293:21
294:1
hearing 5:8 141:16
141:17 157:8
heart 70:8 157:22
165:17 228:11
234:7 260:13
286:9,11 287:12
299:9
heaviest 281:4
heavily 106:16
heck 99:12
hedge 69:1,13
Heidi 2:17 4:16
Helen 2:16 89:3
145:10 289:3
Helen's 208:21
hell 162:9
Hello 112:12
help 96:8 109:17
136:5 154:19
175:14 211:1
242:8
helpful 17:8 26:4
120:10 271:10
helping 4:19
hesitate 279:18
280:14
hesitation 20:20

heterogeneous
294:21
hey 43:7 178:5
high 6:16,17 8:22
10:19 11:13,17,20
12:4,7,9 13:1 17:6
18:17 19:8,16
20:11 21:2,8 23:5
24:6 30:20 31:4
32:17 37:16 38:3
38:6 46:2,3,12
47:8,11 48:5
67:21 91:7 106:22
111:12 114:2,8,13
117:16 123:1
132:14 137:22
145:8 149:19
178:11 190:7
193:6,11 202:8
203:12 225:18,21
226:1,4 243:7,20
253:4 257:18
258:22 266:7
276:20 277:12
279:4,6 285:17
291:21 292:19
294:10,15 298:12
300:2
higher 63:6 66:7
127:18 166:19
185:7 203:15
232:18 262:8,16
278:20
highest 152:8
165:21
highs 30:18 47:16
48:2,6 114:17
high-risk 188:17
188:21 201:13,16
202:19 203:15
209:17 211:5,9
high-sensitive
11:15 25:15,16
high-sensitivity 9:3
23:10 24:9
historical 74:21
history 35:2 36:14

97:22 98:1 130:17
131:2 250:8
255:10 260:13
268:11 291:9
hit 31:11
HIV-positive 85:10
Holly 26:11
Hollywood 177:3
home 179:6
Homer 1:9
honest 59:8
honestly 8:6 234:10
hook 291:2 292:13
hope 52:5 107:21
267:11
hopefully 53:10
59:8 155:13 222:2
227:1
horizon 45:10
hospital 42:21
107:22 108:18
110:5 116:17
127:12 140:4
142:10,13 144:1
150:11,16 155:18
156:14 278:17,22
279:4 297:13,16
297:20 299:1,5
hospitals 28:13
69:12 127:19
128:1,2 137:5,9
140:7 142:6 150:2
155:2 156:5,8
159:2,5 296:4
hour 4:14,15
267:22
hours 146:5
house 69:20 283:10
Houston 166:2
Howard 1:17 26:10
26:12
Howie 32:4 38:10
49:2,7 60:16
65:13 67:3 69:7
70:16
Howie's 35:9
huge 45:9 52:21

Neal R. Gross & Co., Inc.
202-234-4433

103:11 125:6
138:20 172:17
262:1 290:17
hundred 136:8
237:11
hurdle 58:21 102:6
hydrocephalus
131:16 132:10
145:3
hypotension 14:17
hypotensive 13:14
I
Ian 2:18 4:19 5:1
6:5
IB 202:21
ICD 163:22
ICD-9 94:5 99:22
149:9 158:2
228:21
iceberg 263:7
idea 18:22 60:16
75:7 116:22
175:21 213:18
224:16 228:4
238:3 256:5
ideal 29:18,19
146:3
identical 187:17
identification
14:21 246:22
301:18
identified 29:5
99:19 100:15
239:3,21
identify 29:4 41:10
63:10 82:16
100:13 121:2
246:7
identifying 23:19
63:4 255:5
IEP-005-10 3:5
IEP-006-10 3:8
IEP-008-10 3:7
IEP-011-10 3:13
IEP-012-10 3:10
IEP-014-10 3:18


| pre-posed | qualified 259:2 292:1 |
| pre-oped | qualifies 217:4 |
| pre-service 217:12 218:11 | qualify 153:20 254:18 |
| price 32:12 | pulling 60:6 | Q | QI 100:13 |
| primarily 51:1 55:16,18 88:1 | QM 100:13 | Q | QI 100:13 |
| 4(e) 124:1 | 67 301:4,7 |
| 4,000 137:2 | 68 67:22 |
| 4.2 190:20 | 7 3:5 4:7 144:9 |
| 4:00 145:18,20 | 7,000 164:16 |
| 4:40 302:18 | 70 262:19 |
| 40 165:22 209:8 | 70,271 135:14 |
| 281:5 301:11 | 75 52:3 127:14 |
| 400 138:14 139:21 | 141:10 165:11 |
| 414 165:12,15,16 | 750 31:2 |
| 427.61 163:12 | 78464 170:1 |
| 45 3:6 128:6 129:4 | 789,000 150:6 |
| 136:9 137:2 140:5 | 8 |
| 143:11 144:3,9 | 84:7 49:1,2,4,4 |
| 190:21 191:7 | 81:10 92:13 |
| 278:13 | 80 3:8 103:11,14 |
| 450 300:9 | 130:14 155:3 |
| 460 138:15,16 | 80,000 135:8 |
| 462 137:6 | 85 155:4 178:6,7 |
| 48 3:7 | 281:11 |
| 5 54:7 5:18 7:6,11 | 89 3:9 |
| 51:1 94:14 125:8 | 9 |
| 125:9 133:10 | 164:1:189:19 |
| 136:10 202:14 | 9:00 1:10 |
| 300:12 | 9:23 4:2 |
| 5.8 83:6 | 9:30 4:10 |
| 50 30:2 105:4,5 | 90 32:8,11 127:10 |
| 166:20 | 128:11 129:18 |
| 50,000 31:1 | 191:19 192:4 |
| 500 9:17 | 9301 170:6 |
| 6 6 4:7 48:15 49:3 | 95 166:9 |
| 81:10,17 82:1 | 98 82:21 301:2 |
| 203:7 | |