NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
February 18, 2014
1:00 p.m. ET

Operator: Welcome to today's conference call. Please note today's call is being recorded. Please standby.

Lindsey Tighe: Hi, everyone, this is Lindsey Tighe, I'm the senior project manager for the endocrine product. I'm here with Karen Johnson who is the senior director for the project. We would like to welcome you all to our fourth and final workgroup call in advance of our in-person meeting which takes place at the end of next week. Hard to believe it's that soon.

But anyway, I want to thank you all for coming and also thank our developer colleagues for coming. I will go ahead and quickly see who is on the line. Anyone from the Steering Committee, if you could just speak up and introduce yourself briefly, please.

Anne Leddy: I am Anne Leddy. I was nominated by the American Association of Clinical Endocrinologists. So I've been a clinical endocrinologist for over 38 years. And I'm really honored to be part of this group.

Lindsey Tighe: OK, thank you, Anne.

Robert Bailey: Hi, Bob Bailey from Janssen Scientific Affairs. I work in the Health Economics and Outcomes Research team.

Lindsey Tighe: Hi, Bob. Are there other committee members on the line?

Grace Lee: This is Grace Lee from Virginia Mason Medical Center.

Lindsey Tighe: Hi, Grace.

R James Dudl: Jim Dudl from Kaiser Permanente.

Lindsey Tighe: Hi, Jim. Do we have Dr. Rosenzweig on the line? Or Dr. Haydon-Greatting?

Starlin Haydon-Greatting: Yes. Can you hear me?

Lindsey Tighe: Yes.

Starlin Haydon-Greatting: Yes. I'm Starlin Haydon-Greatting. I was appointed by the Pharmacy Equality Assurance group. And I've been a pharmacoepidemiologist for 33 years. I have passed then the director of quality assurance for Illinois Medicaid. I now work in the private sector where we work on employer programs that help with diabetes and cardiovascular at the work site. And I am deeply involved in lots of other quality measure development in all of the other part of pharmacy world.

So, I appreciate the – I was late and something onboard with everybody. There was some – I was out of the country when the appointment information came. I apologize for that but I got things together. But I am honored to be part of this program and be your first pharmacist.

Lindsey Tighe: OK. We're glad to have you on. Any other committee members? All right, great, well then I will …

James Rosenzweig: Hello.

Lindsey Tighe: Hello? Who is that?

James Rosenzweig: It's Jimmy Rosenzweig.

Lindsey Tighe: Great, glad to have you. All right, with that, I'll go ahead and turn it over to Karen Johnson who will give you a little bit more information on how to walk through the measures that we're reviewing in this workgroup.

Karen Johnson: Hello everybody. My name is Karen Johnson. I don't think I've had a chance to talk to most of you so far. But I am the senior director on the project. We had to do a little bit of reshuffling internally. So Reva is no longer working full-time on this project. But she is still alive and well, and so you will be seeing her next week in our in-person meeting. In the meantime you have me to work through the workgroup call today.

So, just to remind you of what we have asked you to do. We have asked several of you to be our primary or lead discussant for the measures. So as you do that, if you would just very briefly describe the measure and also briefly − hopefully you had a chance to look at some of the comments that came through when you did your pre-workgroup evaluation, we asked you to make comments.

So as you perform as lead discussant role, if you would just kind of summarize the comments that were made and just kind of give a flavor of any – particularly any questions or concerns about the measure. So this is our time to really air any questions about how the measure is specified, how to go through the evaluation criteria and the various algorithms, the developers of the measure are on the line with us. So it's also your chance to just directly address them and ask them if there's something in their submission that you don't understand. Now is the time to ask them specific questions.

So, hopefully that's clear and hopefully it will get more clear as we go through the various measures.

Those of you who are new doing this, it may seem a little daunting at first but after we go through a few measures in the workgroup call and especially in the in-person meeting, you'll get the hang of it. So, if you're feeling a little bit unsure, don't worry about it. We'll learn as we go.

So with that, I'm going to hand it off to Dr. Bailey for Measure 0545, the adherence to statins.

Robert Bailey: Great, thanks. And because this is the first time, if I'm not going in enough detail, if I leave something unsaid or uncovered, please feel free to reorient me or ask for some additional information if that's OK.

Karen Johnson: Perfect. Thank you.

Robert Bailey: So the 0545, the measure title is Adherence to Statins for Individuals with Diabetes. And the measure steward is CMS. And the brief description here is that the proportion of individuals with greater than or equal to two prescriptions for statin and have a proportion of days covered of greater than 0.8 during 12 consecutive months.

And so this is specifically focusing on the diabetic population. And the numerator here is individuals who has greater than or equal to two prescriptions for statins for the PDC is greater than 0.8 for statins during a 12-month period. The denominator is age greater than or equal to 18 years of age at the beginning of the measurement period. This is a change from previously as I recall, it was greater than or 18 years of age at the end of the measurement period. So this reflects a change here.

And again, it's greater than or equal to two statins in 12 consecutive months. And the exclusions that are specified here are either a diagnosis of polycystic ovaries and no visits with a face-to-face encounter with a diabetes diagnosis or individuals with gestational diabetes or steroid-induced diabetes with no face-to-face encounter with a diabetes diagnosis. So I guess any questions or comments before we go on to some of the other specifics?

So then transitioning on, this is the process measure here. And using the framework for evaluating the evidence, the pure – the overall evidence in itself is low. There is some indirect link to in either immediate outcome specifically LDL cholesterol and specific threshold. And in the evidence review, several guidelines resided specifically to the American Diabetes Association Guidelines in American Association of Clinical Endocrinologists that recommends the administration of statin to achieve certain threshold targets for LDL cholesterol.

So there's a tangible link in terms of adherence to a statin medication to achieve an LDL cholesterol. But again, adherence to statins is not directly addressed in the guidelines themselves. And then, with respect to the evidence that was presented in the – in the submission, it's looking at various

measures patient adherence but not specifically PDC or proportion of days covered within all of the evidence for such things as the medication cessation ratio, continuous medication gaps but each one of these measures is attempting to get to the same thing specifically medication adherence.

So then when we go on to look at the performance gaps that does exist, there is evidence that is presented to suggest that there is a performance gap first of all. And second of all, that significant disparities exist by when it's – when the data is categorized by age, race or dual eligibility status suggesting again that there is a performance gap first of all. And second of all, that there seemed to be some disparities in certain subgroups as well.

So any questions before I move on?

Karen Johnson: And Dr. Bailey, this is Karen from NQF, my apologies. If it's OK with you, maybe we can go back to evidence and discuss evidence a little bit and then we'd go back and do the individual other criteria. I'm sorry, I didn't mention that earlier.

Robert Bailey: Sounds good. Is there anything in particular that you wanted to call out or you wanted to call out or you wanted to facilitate the discussion on?

Karen Johnson: Well yes. I can pick it off and then maybe the other members can kind of toss to it around a little bit. He said, there seems to be little evidence because the guidelines don't directly address adherence to statins. But the developers did include some other studies. So if we're looking for quantity equality and consistency as a body of evidence showing the link between the measure focus in this case, adherence to statins and the various health outcomes, I wanted to just see what people thought about what was presented in those other studies.

Did, you know, did that seem like that was a review of the body of evidence? Did it show the link between adherence to statins and these prior outcomes? And how do you, you know – can you use that information to go through the algorithm and actually be able to rate quantity, quality and consistency? So kind of toss that three different questions there.

Robert Bailey: And I guess if I could just add one additional consideration with one of my comments as well is that in the diabetes family of measures, we already have a intermediate outcome specifically LDL cholesterol with less than 100 milligrams per deciliter. And now we have a process measure, too, (inaudible) adherence to medications that attempt to get to that part yet.

Karen Johnson: Great point.

Starlin Haydon-Greatting: This is Starlin. Are we to be affected by any of the changes with JNC 8 and the controversy that's going around some of the – now that we don't have a target point for cholesterol?

Karen Johnson: My understanding is that this measure will not be affected by the JNC 8. This is – this measure isn't relying on any particular level of cholesterol before statins are …

Male: The issue is not the JNC 8. It's the American Heart Association, American College of Cardiology.

Karen Johnson: Yes, right.

Male: The guidelines that came out in November which no longer specify targets. They only specify specific amounts of medication you're supposed to use.

Starlin Haydon-Greatting: Right. That's what I mean, Doctor. I grouped them all together because I looked at the whole effect of that gambit that happened in November. And physicians in my community are strategically taking, you know, each group is taking that into consideration for their patient population. But I'm just wondering what the groups – if that affects since we're not supposed to be targeting − is this − are we still supposed to be looking at an LDL-C of less than 100 milligrams in the patient population with diabetes?

R James Dudl: This is Jim Dudl. I think NCQA has put out to retire the 100 milligrams. So I think and the guidelines are out − I think we do have to consider that the nation may switch to being on a statin rather than hitting a target.

Starlin Haydon-Greatting: Right. So, do we address that here?

Anne Leddy: Well, this is Anne Leddy speaking. This measure addresses only whether or not the patient is taking the statin. There's nowhere in this measure has any discussion of the level …

Starlin Haydon-Greatting: It is part of the evidence that's why I brought it up.

Anne Leddy: OK.

R James Dudl: This is Jim again. You know, in the ATP – no, in the one large study, there was evidence that if you take full adherence, you got a 33 percent draw in CVDs to Simva 40. And if you were less – less adherent, it was 27 percent. So there is some adherence data out there, it's sort of indirect, it's not – I don't think this is meta-analysis but I think that's − maybe called incidental.

Karen Johnson: So maybe we should probe just a little bit more about the new guidelines and the evidence would – would that change your feeling about the evidence for this measure?

Starlin Haydon-Greatting: I just think we should probably point to − this is Starlin − point more toward the evidence of adherence rather than it might trip somebody up since NCQA is going along and not − or not using that target that we – we might not want that quoted in our evidence portion. It just might raise a flag to somebody.

Robert Bailey: But what it's sounding like here is that in the absence of a agreed upon target threshold for LDL cholesterol just may be the next step option in terms …

Starlin Haydon-Greatting: Right.

Robert Bailey: … of measuring adherence.

Starlin Haydon-Greatting: Right, right. I mean we have to have – I mean those who are still out there looking at outcomes data, I have to have a number or when I go to the data I can't see if somebody has met their goal or not. So I'm – in the real world I need a number, in the world of trying to create a measure, I am beholded to the experts who are wanting people to be adherent so they're

adherent on the medication and not particularly older patients moving them to an exact number. So I guess I thought I …

R James Dudl: Yes, this is Jim again. We actually – several years ago went to looking at adherence of statins rather than at target and actually did an observational study that showed that there was a difference between those who were high adherent – low adherent. So we have been following that rather than other measures for several years. It works very well.

Starlin Haydon-Greatting: Did you publish that − what was that published observational study?

R James Dudl: Yes, we did it was in 2009 in a channel that I can't remember right now …

Starlin Haydon-Greatting: I mean − but that's a good example.

R James Dudl: Yes, it is. I'll get it to you, I'll e-mail you.

Starlin Haydon-Greatting: Yes, that would be great. I think that would add something and …

R James Dudl: Again, I think there's indirect evidence, this observational type of thing but it was there.

Starlin Haydon-Greatting: All right.

Robert Bailey: And –this is Bob Bailey. And I think the other measure consideration here, you know, understanding that there are limitations, I guess the issue is when you go back to a process measure, it's very much like going back to hemoglobin A1C monitoring if – if the test has been done but not only the test being done but is the value within the target range and here, where you're looking at PDC, you're not necessarily determining whether the appropriate dose of statin is being used or just determining whether the patient is being administered a statin and whether if they're adhering to the dose that's prescribed.

Starlin Haydon-Greatting: Agreed.

Male: I would just say that there's a lot of controversy about this.

Male: Yes, very much.

Male: The American Association of Clinical Endocrinologists came out strongly against it and several other major organizations are probably going to come out against as well.

Male: I don't think we're recommending retiring the others or recommending adding this.

Female: Yes.

Male: And I would also mention that, you know, with respect to the medication adherence measures, the three of them that are evaluated today. This is probably one where the evidence is probably the strongest in terms of adherences and association with an outcome?

Female: Yes, yes, yes, great. I will just – I just thought we should probably discuss that at this point because of the controversy just to make sure that it was – added on the table. And I was interested in all your opinions because I respect you.

Male: Opinions about?

Female: About how that – that being stated in the evidence what changes happened in November and how this would impact this measure at all, that's what I was interested in, because I, too, am doing an observational study in our patient population on adherence to statins, so.

Robert Bailey: Well, it's − I find it, yes, it's really an opinion on my part, I can't – I don't want to go against the whole cardiological community but – but it seems that the way that they're set up, they're very difficult to implement appropriately because you're not treating to a target, you're just giving a certain dose of medications based upon …

Female: Right.

Robert Bailey: Actually, I'm not even at cholesterol level but rather risk level.

Female: Yes.

Robert Bailey: And the risk engine has some problems with it, the one that they use and in addition, there is a sense that you're supposed to treat to let's say, you know, you're using high dose statins to have more than 50 percent of your original cholesterol level but you don't − if you're inheriting a patient from someone else, you don't know what the original level was. So it's very hard to implement successfully, however, I'm trying to use it in some of my patients. So I'm not − I don't want to come out of the ACC/AHA measure but I think a number of other organizations are going to find it difficult to adhere.

R James Dudl: This is Jim. To my knowledge is − when you propose a change this big to a nation, it takes three years for them to absorb. I think we're going to need both for a while and then in three years, we will know which wins out.

Robert Bailey: I agree with you.

Female: OK, so good, then we can − since it's − actually, they were stating that it might take ten years before anybody even we start seeing the patients. So this has been implemented in before we start seeing the outcome and information because it's going to take that long to change people's behavior − professional, the family practice physician from the internist and the cardiologist have their processes in place. But it's going to take 10 years before you go to see what the impact of this if is and then they'll come up with something different by then anyway so.

R James Dudl: By the way I found that reference in my mind again it's …

Female: Oh, good.

R James Dudl: Look up Dudl, D-U-D-L in American Journal of Managed Care, it usually pops up its own (line).

Female: OK, thank you.

Karen Johnson: OK, so this is Karen again and (inaudible) dialogue for me to listen to but in terms of this measure in the evidence for it, do you have any questions or any

− do you have any flavor about − do you consider this to be a systematic review? Are you able to actually apply the evidence algorithm and arrive at a rating? And would it useful to walk through that algorithm?

Starlin Haydon-Greatting: Yes. I think it would be very helpful.

Karen Johnson: OK.

Starlin Haydon-Greatting: Yes.

Karen Johnson: Lindsey can you turn that up? Lindsey is going to bring up the algorithm for us.

Robert Bailey: And I guess while she's doing that, these seven studies I guess in my mind doesn't necessarily equate to a systematic review of the literature.

Male: Agreed.

Karen Johnson: OK. OK, so Lindsey has the algorithm up and bear with me, I'll just kind of walk you through the first − the green bars are there and those apply to outcome measures. So we can ignore boxes one and two because we're not talking about outcome measures, we're talking about a process measures. And then that puts us to box three and the first question is, is it based on a systematic review and grading of the body of evidence where the specific focus to the evidence matches what's being measured?

So this could come in box three. It can come from a variety of things. So it can come from systematic review, say, from the (Cochran) collaboration or, you know, any other kind of groups that might do a systematic review.

It can also come from systematic reviews that happen (within) critical practice guidelines are (Korea). So, the developer did offer a critical practice guideline, so I guess one of the questions are, did you feel like these guidelines did or did not actually basically apply to the measure as specified? If not, then we need to think about the seven studies that were described for us. So I think that's the first question that we want to answer.

Robert Bailey: And I can take a stab at that because the CMS does acknowledge in both of the guidelines that are cited here that the guideline do not address the topic of medication adherence directly. So then we − I'd recommend we have to default to the seven additional studies.

Karen Johnson: OK. All right, anybody on the committee disagree with that one? OK, so then that takes us I think to − can you go down just a little bit, Lindsey.

OK. All right, so pretty much the seven that – so we have kind of a choice here of whether we go to box seven, where box seven talks about there is empirical evidence but not with – through a systematic review. Or we can go back up to box four, I think it is, if you considered the seven studies of systematic review. So do we feel like that these seven studies would be a systematic review?

Robert Bailey: And I would advocate, the answer is no.

Karen Johnson: OK. All right, just for time, we'll just kind of keep walking through. So (that's) in box seven. So, is empirical evidence submitted even without a systematic review? And we would say yes. That is the empirical evidence that is summarized including all the studies in the body of evidence. So I think the developers did describe how they came to the seventh study.

Do you still like that they've got everything? Everything that would be relevant, I think. And this is a trick question. I mean, we will rely on your expertise and your knowledge of the field, obviously, to help us answer the question.

Robert Bailey: You know, I'll take the first step, I would say no, I mean the seven studies they did included were important but I'd advocate that it wasn't necessarily the entire body of the evidence.

Karen Johnson: OK. Any other feelings about those studies that were – if the answer is no then our algorithm will take you to low.

Robert Bailey: And I guess the other consideration here as well and I mentioned this earlier that they all – the studies that were cited look at various measures of adherence but not necessarily directly at proportion of the days covered itself.

Karen Johnson: OK. Any other committee members who want to talk anymore about box eight and whether you would answer yes or no about a box eight.

Female: Probably it'd be a no.

Karen Johnson: OK. All right. So, hopefully that at least gives you a flavor of how you walk through these algorithms and, you know, evidence is one of those – one of the criteria that is sometimes a little trickier to work your way through. And to say you understand what these purple boxes do, we do realize that sometimes there are measures that – maybe there are – is not a systematic review that's been done.

So, you know, the purple boxes in the scenario where there hasn't been a systematic review of a particular scientific question or something. But we still want to make sure that even if a systematic review has not been done that you are able to consider the entire body of evidence, you know, that there's not cherry-picking of things so that, you know, you don't really understand the full body of literature that's out there.

So, depending on how you feel about question eight that, you know, you could rate this as low on evidence or potentially rate as moderate. But the fact that there is not a systematic review does limit − there's no way that this measure I think could get a high rating for evidence. So hopefully that makes some sense to you.

If we were in the in-person meeting, every, you know, everybody in the committee would have a chance to talk through some of these things and then we would ask you to actually vote on evidence. If there was a really strong majority who feel like that the rating of this measure would be low then that would pretty much conclude the discussion of the measure. In other words, we wouldn't take it further and talk about the other criteria. But since this is not the in-person meeting, we will go ahead and continue working our way through this measure.

So with that, let's go head and talk about gap, (inaudible). Dr. Bailey, did you want to go ahead and do the gap for us?

Robert Bailey: Sure, no problem, sorry. So, based on the evidence that's provided here, you know, there is evidence that there is a gap in terms of attainment of that PDC threshold of 0.8, when you look at it from a drug plan or physician group perspective. So there seems to be enough of a gap in care who were in the national performance measure here.

And then when you drill down a little bit further to look specifically in terms of disparities, there does appear to be evidence to suggest that there are significant disparities both from an age and race perspective, but also from a dual eligibility perspective as well. Another question that's posed, should this measure be stratified for disparities? And to address that question, I would propose that that may make sense from several perspectives. The first one is identifying potential targeted opportunities to improve the disparities that are identified here.

And then in terms of priority, you know, based on, first of all, the high prevalence of diabetes in the U.S. population but also the high cost in the associate (heights) health care resource utilization, it does suggest that this is a priority condition.

And – sorry, go ahead.

Karen Johnson: That's great. Does anybody on the committee have anything you want to add to that synopsis of disparities and priorities?

So generally, since we have a limited time, we will limit our discussion for the most part of things where there might be disagreements or maybe some non-clarity. I will ask Lindsey just to make sure everybody is aware that the gap and care and the priority sub-criterion, for that when you would use – we don't have a fancy algorithm for that, we just have our generic scale.

So based on – you basically look at what is presented to you and just write down your confidence or certainty that the criterion has been met. So a little bit easier scale to have to deal with, I think.

OK. Dr. Bailey, why don't we go ahead and go to the liability and go ahead and start with the specifications and then we'll talk about some (inaudible).

Robert Bailey: That's sounds good. So from the specifications, they appear to be clearly defined and to address the question in terms of could they be consistently implemented, I would advocate that their – the definitions are clear enough that that should be the case. And with respect to the reliability testing is meet the predefined threshold of greater than 0.7.

And there was a question that came up in terms of the validity of the specification and to address the question whether it is consistent with the evidence, it's not entirely consistent with the evidence however it's logical to assume that the populations that are identified in the numerator and denominator inclusions are, you know, are lined closely enough to include the population of interest.

And with respect to the validity testing, two potential threats to validity were identified specifically missing day supply and cash prescriptions and those were appropriately addressed in terms of sensitivity analysis suggest that they – it would not pose significant threats of validity here.

And then the other question was posed is in terms of the test sample was that sufficient to be generalizable. And although the test sample was limited to a certain number of states, not necessarily completely geographically represented of the United States population, there's no reason to believe that the applications would be different, that – and that would not be generalizable.

And the other question was, you know, the exclusion of – hold on, with the exclusions, I mentioned this earlier that not completely consistent with the evidence, but it did make sense that the population is being isolated interest. Any other questions that I haven't raised here or any other points?

Karen Johnson: So this is Karen. Just a couple of things I'd like to ask you. The measure is actually specified for several levels of analysis, including health plans and clinician groups, other delivery systems. And it looks like that the developers actually did give a liability testing for those different groups that I just wanted to see if anybody had any concerns about any of those particular levels of analysis.

Do you feel like it is reliable for all these different groups that they specified it for?

R James Dudl: This is Jim. I'm not sure but at 0.9 for ACOs in states, you would think that high reliability for physician groups of 0.7, I think it's moderate. I mean, yes, when it comes to that when I do it. And I just – in my mind I thought that division might be appropriate. But then, if you do that, you give the overall group a moderate because that would include everybody.

Karen Johnson: All right, yes. Yes, pretty much because we won't ask you to vote on each individual level of analysis separately. So you're kind of having to – kind of average out your …

R James Dudl: Yes, OK.

Karen Johnson: … yes, average and not the statistical mathematical way of saying average.

Robert Bailey: And I guess a question along those lines because even for the physician group, it still meets that threshold of, you know, where it is reliable, although there is a significant difference between physician groups and the other groups that are being evaluated. So is that something that you're taking consideration here?

Karen Johnson: Yes. And just so we're all very clear that the 0.7 is kind of a rule of thumb for reliability. There's some people that think that 0.7 really isn't high enough when it comes to maybe looking at individuals and comparing individuals but for the most part 0.7, you know, is pretty generally the rule of thumb. So, you know, you kind of have to take that in mind.

I think the question that I had and I wasn't completely sure as I was looking at some of the numbers here and maybe this is a question for the developer. There is several physician groups particularly or even the drug plans, there is generally one reliability for different years that are given. And what you would see when you're doing this analysis, you'll actually have a reliability score for each health plan or for each physician that is being considered.

So I was curious, is this an average reliability score that's being shown here, is that what is in the submission? And that's – it's just a minor question, but it does look like results are pretty consistent.

(Off-mike)

Karen Johnson: Yes.

Female: You know how to answer that.

Karen Johnson: Do we not have a developer on the line that could – right, it's really a minor point and we won't spend any more time on this. As I see the results for the ACOs, I, you know, that's exactly what the physician …

Kyle Campbell: Hello.

Karen Johnson: Hello.

Kyle Campbell: Hello. I am sorry my line apparently was muted, so I just have the operator unmute my line. This is Kyle Campbell and I'm the executive director of the project for CMS at FMQAI, also a pharmacist.

And just to answer your question, yes, in these cases since we had so many entities to present, we did just present a mean reliability score of points where the threshold was with regard to 0.7 for a minimum denominator, and those minimum denominators that you can see in the table, for example, for the plans are 260 and 200.

If needed, we can present a full table similar to what we presented for the ACOs.

Karen Johnson: OK. Right, so let me just make sure just for everybody that I understand Kyle and thank you for that. You're saying that to get a reliability score of roughly 0.7 for the statin measure for your drug plans since 2012, you need at least 200 patients?

Kyle Campbell: Correct, 200 patients within the denominator and then for physician group, it's 175 patients.

Karen Johnson: OK. All right. Just for future reference, another – it is actually that you shown the denominators because we do understand what these reliability scores are a function of sample size, so the number of patients who, you know, that's really good stuff to know.

In the future, you may want to consider just doing a distribution of your reliability scores as well that would, you know, you certainly wouldn't have to do like what you did for ACOs, because there's going to be a lot of physician groups and health plans and stuff, but …

Kyle Campbell: Sure.

Karen Johnson: … that would be just another option to …

Kyle Campbell: To consider, sure. Thank you.

Karen Johnson: Yes. So the committee, now that Kyle is on and can actually be heard, do you have any questions for him about anything that they've done in terms of their reliability testing or their validity testing or any questions about what they looked at in terms of their threats to validity? Do you feel like that they adequately addressed the cash prescription question, for example, and did anything else come to your mind that you may want to ask them?

Robert Bailey: You know, I guess I'll go first, if I may, you know, from the cash prescription, it seems as though they answered that question. And I guess the only other potential threat would be just to have a patient that has prescriptions for more than one type of statin, you know, because they're either changing therapy or they were intolerant of a particular therapy, you know, how does that enter the consideration here?

Kyle Campbell: OK, so the answer to that question is the rate is capsulated across the drug class. So if the patient switches statins that's accounted for in the PDC algorithm and they're credited for that, you know, any algorithm (inaudible).

Were you able to hear the response OK?

Robert Bailey: Yes, thanks, Kyle.

Kyle Campbell: OK, yes, sure. I was just concerned maybe that my line had gotten muted again. Sorry.

Karen Johnson: There is some interruption.

Committee members, this is Karen again, would it be helpful to go through the reliability algorithms or are you guys pretty comfortable with how that (inaudible)?

Female: I would certainly like to go through it.

Karen Johnson: OK. Lindsey is bringing that up now. And I know we have a lot more to get through, but since the other two CMS were just so similar to this, I think this is probably good use of our time to make sure that we can get through the algorithm.

OK, so for evaluating reliability, the first question we have to answer is about the specification. So, again, specifications having consistent and precise unambiguous specifications are the foundation for having a measure that can be consistently applied and implemented. So that's why there's that first star there. And it sounds like everybody felt like the specs were fine. So that takes us to box two, was empirical reliability testing conducted using statistical test with a measure is specified? And, I believe the answer to that one was yes. They actually computed with their percentage of days covered.

And then that takes us to box four, with reliability testing conducted with the computed performance measure score for each measured entity. And the answer of that one is yes. So they did the signal-to-noise analysis and they did it for the various levels of analysis that they specified, so again the health plan

the group's mistake, all these different things. So then box five is, was the method described inappropriate? And again they used the signal-to-noise analysis. And there's a tutorial that tells us how that works. And that is definitely an adequate method, inappropriate method. So we would say yes there.

So that takes us to box six. And this is where you really have to start thinking about the numbers that came through. So you look at two things to answer box six, you look at the sampling that was used. So, in other words, you know, what sample was used to test. And in this case, I believe it was data from or the 10 – 8 to 10 states. So, I believe there was information about how many patients that included.

So you think about the scope of testing and then the actual results of the testing. And then you decide whether you feel that there is high confidence or certainty that the scores are reliable or moderate certainty or low certainty. And what you'll see here for the algorithm is – because they tested at the score level, there is a potential that they could get a high rating for reliability. If we went further down the algorithm, if they had only tested at the data element level then the highest possible rating would have been a moderate level.

So, again, you pretty much you think about what level of testing was done. And then you think about the methodology that was used, the testing sample that was used and if adequacy is generalizability and then finally you think about the results. And you put all of those things together and decide on your way. So, let me stop here and see if there's any questions about going through the reliability algorithm.

Hopefully, the algorithms are clear. I will tell you as I've told other folks on the committee, this is the first time we've used these algorithms so we'll be very interested in hearing your opinions in, you know, are they understandable, can you consistently apply them, that sort of thing. So …

Female: I thought the reliability algorithm was really clear. I just marched straight through it and got to the same conclusion. I'm doing the one for adherence on oral hypoglycemic. So, I thought it was really well done.

Karen Johnson: OK, good.

Robert Bailey: I agree, I thought it was well done, it was very helpful.

Karen Johnson: Good, good. We won't look at it in detail unless we need to, but the algorithm for validity is similar. It's structured in the same way. It is a little bit more complex because as we've already talked about with validity, it's not just thinking about how the evidence matches up with the specs and what kind of testing was done and what those results were, but you also have to think about attempted threats to validity. And you have to kind of work back into your rating as well.

With these measures, it wasn't – it's not completely complex, there were exclusions. They would say talk about − but sometimes if you were doing, for example, an outcome measure that have risk adjustment, you have to incorporate all that thinking in critique of a risk adjustment model when you're thinking about validity. So, it can – sometimes can be a little bit more complex than what you have in front of you for this measure.

OK, let's go ahead because our time is really scooting by. Dr. Bailey, can you walk us through feasibility.

Robert Bailey: Sure. And so from a feasibility perspective, you know, the question was posed, you know, is this data routinely generated during the routine care delivery? And the answer here is yes because it's based on pharmacy claims. And so I don't see any challenges there. And then the question comes up use and usability, and this is not currently used for public reporting but the planned uses are public reporting and quality improvement.

Anything else that I haven't addressed?

Karen Johnson: Do you – and I don't remember if this was a question specifically, is there any concerns about initially the part of the usability in this criterion. Is there any concerns about any potential or unintended consequences in this measure, is that anything – did anything come to your mind that you want to air?

Robert Bailey: Nothing came to mind in my review. But I think the only concern that was raised from my perspective, I'd be interested in others as well is the timeliness of the data back to these various groups that are being measured. So first of all, they can understand where they are and more importantly can − understand or study the impact the interventions that they're making.

Karen Johnson: Is that something you'd like to address to Kyle?

Robert Bailey: Sure.

Kyle Campbell: So the measurement period is for the full measurement year. And the claims data typically do lag about six months. You know, in our experience with regard to interventions related to patient adherence, we think if, you know, if practices were reviewing data potentially stratified data or ACOs to look at their overall level of performance that these data would, you know, potentially be timely enough for them in terms of intervention.

You know, typically, it will take a while, maybe two or three physicians with a patient in order to get them back to an adherence data. So, you know, that would show up in the following, you know, measurement a year, the way the measure is structured.

Robert Bailey: OK, thanks.

Karen Johnson: OK. And since this is the workgroup call, we're not going to really talk much about that related and competing measures. I will tell you if you can see on your staff review that (inaudible) measures and even the developer has identified several related measures.

So, in the in-person meeting, we will be talking about these and discuss, walking through a little bit to make sure that these related measures are harmonized to the extent that they can be. And the point of that exercise is it can be confusing and difficult for implementers of the measures to have to do measures that are similar but not quite the same. For example, defining their diabetes population. So that's why we do this related and competing measures discussion. So, again, we will leave that for the in-person meeting.

OK. Is there anything else burning that we would have talked about with measure 0545 or are we ready to go the next one which is 2467?

Kyle Campbell: Karen or Lindsey, this is Kyle here. Could I ask a question?

Karen Johnson: Sure.

Kyle Campbell: So I think based on the discussion around the evidence that the evidence review form might not be the latest version that was submitted to NFQ for 545, because we did cite the systematic review from the cholesterol trials in 2010. And then we had – in addition to that, three additional studies that directly relate to adherence.

And in terms of the evidence for the 2013 ACC/AHA guidelines, the first recommendation with regard to diabetes as moderate intensity statin therapy should be initiated or continued for adults 40 to 75 years of age and that the class 1A strong recommendation.

Karen Johnson: OK. And I think Lindsey is taking a peek right now. And we will – I'm looking at it, too, Kyle.

Male: Yes. I thought there was much more strong evidence out there. So I think that's very worthwhile.

Karen Johnson: OK.

Female: Yes, I wondered why the 2010 didn't show up on the evidence because that's – that's the evidence we used part for the PQA measure.

Karen Johnson: Of statins.

Female: Yes. Thank you Kyle for bringing that to our attention.

Kyle Campbell: Sure, yes. I just wanted to make sure in the evidence review that you have the latest evidence to make your consideration.

Karen Johnson: I don't know if we have a date. We'll have to look and make sure because we do want to make sure we have the latest thing. And the only other thing that I

can think that might be is – we did the staff review, so we did them fairly early on.

So, Kyle, do you know if you submitted something a little bit later? Or …

Kyle Campbell: What happened is NFQ staff had recommended that we list the measures …

Karen Johnson: Right.

Kyle Campbell: … you know, from the original NFQ 545.

Karen Johnson: Right.

Kyle Campbell: And so, when we split the measures that necessitated, you know, and allowed us to have additional space to put in evidence, you know, specific to the drug classes.

Karen Johnson: OK.

Kyle Campbell: So it's possible that what the committee has is – was the original submission for evidence prior to the splitting. And then the data submission for the most recent one would have been 1/31/2014.

Karen Johnson: OK. We will take a look at that and get back to you and make sure.

Kyle Campbell: OK.

Karen Johnson: Yes.

Lindsey Tighe: Yes, I think we did provide the correct attachment to this committee. It may just be that the staff review that on top did not have the links …

Karen Johnson: Yes.

Lindsey Tighe: … provided so we will update that.

Karen Johnson: OK. So committee members, be on the look out from us. We will take a piece of that and make sure …

Female: Will you indicate that on the PDF that is the updated one when put it in the SharePoint?

Lindsey Tighe: Yes. We only keep the most updated version of the PDF on SharePoint, so there won't be competing version to confuse you.

Karen Johnson: Yes. We made just for clarity. We were actually talking about this, this morning. We may do something like a track changes just to make it very obvious.

Female: Oh, that would be wonderful.

Karen Johnson: Yes. I know for me, I'm old. I print things out. So, it's kind of hard to tell if I got the most recent print out, you know. So, we'll definitely get back to you but thank you Kyle for bringing that to my discussion since the committee members are very familiar with what you're speaking of, so.

Kyle Campbell: Sure. Thank you.

Karen Johnson: With that, I think it's still probably useful to walk through measure on 2467. As Kyle said – and just so everybody is clear about what happened. When this measure was originally endorsed by NFQ, it came through with – my understanding is that it came through with three separate scores all kind of classed into one measure. And that's not how we do things at NFQ anymore. We like to have one score per measure. So we actually asked the developers to split them out.

So now, instead of having one measure with three scores, they have three measures with one score each. So, they did that for us. But the minor say is the methodology and some of the evidence at least it's going to be similar for the two additional measures. Well, maybe not so much the evidence but I would imagine your testing methodology is actually the same.

So let’s see trying to see who is our primary discussant, Dr. Dudl, you're our primary discussant for 2467.

R James Dudl: Thank you (for that).

Karen Johnson: Yes.

R James Dudl: First of all, I can't thank you enough for making me second instead of first. All of the heavy lifting has been done and so we could – I'm going to really be able to just defer to all the specifications that were stated for most of this instead of repeating them and we'll go a lot quicker.

I will briefly describe the measures addressing adherence to ACE inhibitor, ARBs in proportion day covered at least 0.8. The rationale is to identify individuals with diabetes who are not adherent with the rationale that ACE − ARB adherence is expected to result in lower rates validated blood pressure, cardiovascular events and mortality.

So, with that, more brief, I think we can go right on to looking at the evidence which does pose some challenges. One second here when I get to what I wrote, yes. So, first of all, it's process measures so the quality is appropriate for that.

And it was mentioned, this is not looking at all agents for blood pressure control because they're really looking at cardiovascular benefits from the ACE rather than blood pressure control from the ACE. And they do cite good systematic reviews of ACE in CVD with diabetes and support for trials and even better than the JNC 8 data review of the effects of ACE and ARBs on CVD finding moderate evidence for starting and ACE ARB is one of the several choices.

So I think the evidence hat ACE ARBs work in preventing CVDs in diabetics is solid. But the question was is there evidence of adherence being a significant metric? And the answer was that at least, you know, the data that I got was it wasn't reported here. And that their support was that the ADA, AACE did say that they are using it.

And I don't know if Kyle is there – that we got than what I just reviewed?

Kyle Campbell: Yes. So, I'm going to have (Liz Sloss) who's on the line from RAND speak to that. (Liz), you have to ask the operator to unmute your line.

R James Dudl: (Liz) when you're unmuted just speak right up and I'll keep talking until you're (inaudible).

(Elizabeth Sloss): Hello.

R James Dudl: Is that (Liz)?

(Elizabeth Sloss): Oh, I don't know. Can you hear me now?

R James Dudl: Is that (Liz)? Just say, can you hear me now? Yes.

(Elizabeth Sloss): Yes, yes, yes. I am not sure, given what you said Dr. Dudl. You may have the latest set of evidence. The evidence that we presented was from the Blood Pressure Lowering Treatment Trialists' Collaboration published in 2005. And is that what you were referring to? There were four studies within that meta-analysis that we …

R James Dudl: Right, yes.

(Elizabeth Sloss): Yes. OK.

R James Dudl: I got that one, right.

(Elizabeth Sloss): Yes. OK. So that's the bulk of the evidence that we reported in the forum.

R James Dudl: OK, great. Let me move on then. The others, you know, I knew this. So I do – and I thought I read it but you allowed a review of the grade criteria for implementation quality, benefit, outlaying risk, costs, patient preferences as being important factors besides quality of evidence. Is that true or not? I think …

Could you hear the question − anybody on this test?

Kyle Campbell: (Liz), are you there?

(Elizabeth Sloss): I believe that Dr. Dudl − were you addressing that to NQF, Dr. Dudl?

R James Dudl: NQF, yes.

(Elizabeth Sloss): Yes.

R James Dudl: Thank you. Sorry. So anybody from NQF to just grade criteria, is that something that would we could quote as a valid reason for implementing rather than evidence alone? Well, let me do this, OK. I will say that we use grades and grade is, you know, in the literature, and I'm going to quote how this might fit within grade.

So, first of all, even though there wasn't any evidence that was quoted for adherence, there is empirical evidence in trials linking heart disease. One article and I quoted American Heart Journal showed CVD that changes with 80 percent adherence to ARBs where 15 versus 20 very significant preferences and hospitalizations revascularization. So not – and some of those, 705,000 were diabetics.

Another paper which was done – only diabetics but with ischemic heart disease, so not pure diabetes, or decrease in death with adherence proportion of days covered both of these. So I would think that this would – I don't know if these are incidental pieces of evidence or not but I would suggest searching for incidental or other evidence. And the reason is that, you know, this has been proven to be very effective in diabetics in decreasing cardiovascular disease. And even if the evidence for adherence is weak, the benefits definitely outweigh the harm. It's been modeled to be cost-saving. So side effects are known and manageable. Patient preferences, well, you asked them to, is it worth taking a pill for decreasing heart attack. And so, in our experience it's high.

So I would say we have to call this insufficient evidence right now. But I would wonder if going back and reviewing for incidental might raise it to a grade level where this could be kept and used. Let me stop with that and throw that out to everybody because that's a – and there's a bunch of stretches there and I don't know if this is a group where we do those stretches or where we say, "No, we're just going to stick to the criteria."

Karen Johnson: Well, this is Karen from NQF. And you can, I mean, one of the reasons that we ask for experts of your caliber on our committee is that we know that you

guys know literature in that sort of thing. So you can certainly pull from your own personal experience and knowledge and apply that.

So, yes, I think the answer is yes that you can use the study that you know about, you know – the only thing that's a little tricky is that there are, you know, it's a multi-stakeholder group, and while we do have – many of you who know the literature really well, not everybody will have that level. So, you know, your explanation and your discussion is some of the studies that you know about will be helpful to the other committee members who maybe do not know that literature quite so well.

R James Dudl: Sure. And then I would throw it back to Kyle and RAND whether they thought it would be worth looking for this other data.

Kyle Campbell: So this is Kyle. And then I'll (inaudible) or (Liz) respond from RAND. I think in this case, because adherence really haven't been studied across every drug class for which we would like to have a measure, I think what you proposed makes a lot of sense in that, you know, pulling the trials that have demonstrated clear outcomes for ACE/ARBs are related to patients with diabetes (inaudible) the trials as if the patients who were adherent to the medication with a certain proportion.

Now, whether that proportion was actually 0.8 PDC, you know, that's not entirely clear and those data are not available from the trials but presumably the trials did, you know, have some sort of metric of insuring participants would adhere.

So I think that that is a reasonable, you know, course of direction and also we'd have RAND, (Liz) and or (Soren) jump in there.

R James Dudl: Yes, the 0.8 was – there was 0.8 used. I think its PDC but it wasn't – if it wasn't, it was very close.

Kyle Campbell: And that 0.8 is the standard among adherence measures across the NQF portfolio. And, you know, we have several measures that are adherence measures and the proportion of days covered has been really agreed upon by

most developers with the standard to use to measure medication adherence with administrative claims data.

R James Dudl: So again, I would say at this point, our algorithm says it's sufficient but I think if we would include those and I don't know if you would review those and see if they're includable or if others felt that they were, then I would then put the evidence down as weak and then we could proceed.

And I'd like to that as for now and not – and review the risk because the risk is going to go very fast, and suggest that you consider looking for other incidental that could give us a weak recommendation that maybe usable with grade and that's – OK?

Any comments from the other reviewer or anybody else? I don't want to push this on anybody. I'm new to the game but it's an awfully, you know, it's an awfully good evidence that the drug works.

Starlin Haydon-Greatting: Well, it's good evidence because – and people are – have implemented it into their – I mean even the HEDIS measures are checking to make sure that you have your patients on an ACE or an ARB. So I agree with you. I just think we're ahead of − the studies that support all this are incidental because once we saw that it worked, why go through more, you know, study processes. So I think we're going to have to add that incidental evidence. This is Starlin. Thank you.

Robert Bailey: Thanks.

R James Dudl: Do you want me to continue or do you have any other comments yet? Any other comments?

Karen Johnson: And this is Karen. Just keeping track of time.

R James Dudl: Yes.

Karen Johnson: Are there other pieces of this measure that, you know, either with the reliability testing or some specs or any exclusions? Any questions at all that

you want to bring out and discuss rather than – I don't think we have time to (inaudible).

R James Dudl: They were really pretty much discussed already and all the criteria with all of them. I think the rest of it sails right through. Nothing else that I would comment on.

Female: I agree.

Karen Johnson: OK. OK, so it sounds like in both measures at least so far is that the sticking point has been the evidence. And we'll definitely have to go about that systematic review and make sure that it's very clear to everybody for the first measure.

And then go – potentially had – what we can't do right now is open up the submission for Kyle and his group to add to the submission form. But we will reflect this discussion in our summary notes. We're going to write up the summaries of these workgroup calls since we'll have to − yes.

R James Dudl: OK.

Karen Johnson: Let's go ahead to the third one, adherence to oral diabetes, and I think that's Dr. Lee.

Grace Lee: Hi. So this measure is the third in the series submitted by CMS, Adherence to Oral Diabetes Agents for Individuals with Diabetes. This is a process measure. And then very similar light, this measure, it measures adherence again with the proportionate days covered.

So, basically, is report – percentage of eligible individuals with diabetes who had at least two prescriptions for a single oral diabetes agent or at least two prescriptions for multiple agents within a diabetes stroke class. And PDC is again at 0.8 for at least one diabetes drug class during the measurement period for 12 consecutive months.

So just to double a little further, the numerator is individuals with at least two prescriptions for oral agents and any diabetes drug class with a PDC of at least

0.8 for at least one diabetes drug class. And the denominator is – again individuals at least 18 years of age at the beginning of the measurement period with diabetes and at least two prescriptions for a single oral diabetes patient or at least two prescriptions for multiple agents within a diabetes drug class.

Speeding along, a lot of this measure hangs again on the evidence. And I guess we can open this up to CMS in a little bit. So looking at our algorithm, the CMS did submit that there were practice guidelines suggesting that obviously, oral hypoglycemia must be used but that – the guidelines did not address specifically the topic of medication adherence directly similar to the previous two measures.

And then the developers presented the results, the review of six studies. So – and one of the first questions was, is this a systematic review? And I would argue and also summarizing comments that have come in that this is no, this is not a systematic review.

So they did look at these six studies on adherence. But again not PDC per se and looked at a variety of outcomes both hospitalization, hemoglobin A1C and other measures. But the − there was a systematic review including both adherence to oral and insulin, that was in 2000 just looking at the number of different studies that they covered in that study. It appears that some of the studies that would fall under the criteria for review were not included in those six studies.

So looking at that and looking at the quality and quantity and consistency, I would have to say that perhaps some small studies were left out in which case in the algorithm that puts you at a lower rating. Should I open up for comments at this point?

Karen Johnson: Yes, please.

Grace Lee: So, my one comment to CMS like the other measures is what − were their other evidence submitted that I am unaware of − of studies?

Kyle Campbell: So those studies that are – I believe you have the correct form because there were six additional studies submitted.

Grace Lee: Six additional or six only?

Kyle Campbell: I'm sorry, six only.

Grace Lee: OK.

Male: What was your grade that you gave for quality?

Grace Lee: I was a bit tossed up on this. I think my original notes – I was hovering between a low and moderate because, you know, the sticking point for me was that, that they did include all the studies out there. And there seemed to be a large number of different studies they could have included. But the quality of the studies actually was good. They used large studies. They were from integrated systems. One of them was Kaiser, Colorado, but the VA, they had used outside groups so one study was from Korea.

So, you know, the quality of studies were actually − was actually moderate. And I was (sure) tossed up between moderate and low. But the problem I had was that it just wasn't as comprehensive as I would have liked. And also that there was – because they had just looked oral agents, there wasn't a systematic review on oral agents. However, there was a systematic review on oral and insulin.

R James Dudl: Yes. It was the second review, I gave it moderate. But I totally agree with your other comment about, you know, not including – it wasn't systematic.

Karen Johnson: So this is Karen. Do you know – you mentioned the 2011 systematic review. Is that something that you might find a citation for that perhaps we could share with other committee members?

Grace Lee: Great, I have the PDF. So it's from (Ash et. al.) from clinical therapeutics in 2011. And just because I'm not really an expert on adherence, so I started looking around the data and then found this nice review that encompasses it. I think 37 − (37) percent. And they went through, you know, very clearly how they came to these studies and very clearly laid them out.

So, just looking at that. I was looking at it like, well, you know, there are numbers of different studies that seem to – would be included in a review. And they were seemed to be left out but that's just my (cursory) look.

Karen Johnson: OK. If you wanted to share that with the rest of the committee members, they can certainly push that on our SharePoint site so that other members can take a peek at that as well.

Grace Lee: Sure.

Karen Johnson: OK. If you want to e-mail it to either Lindsey or I, either one.

Grace Lee: OK, great.

Karen Johnson: Were there other – I know that there is tons of other things to talk about with this measure but unfortunately we weren't going to have time. Was there any other thing with the other criteria, anything that you wanted to bring out particularly things that you had questions on or it seemed like maybe were unfair?

Grace Lee: I think that the reliability and validity testing was very similar to what had been discussed. They have very nicely did cover potential threats to validity specifically those cash prescriptions. They did a very nice job of that.

And one of the other comments was – and I guess we're going to talk about competing measures, but one of the comments that was posted was that there is – this is an interim outcome. An interim outcome measure does already exist specifically glycemic control as manifest by hemoglobin A1C. But I suppose we are going to talk about that more in person.

Karen Johnson: Yes. Yes. And one of the things also that we'll ask you to do as the standing committee, you guys are kind of responsible for our portfolio of measures. So, you know, you're kind of the owners and the guardians, if you will, of the measure. So those kinds of things come up and it's kind of the over arching question even beyond the related competing measures.

Robert Bailey: Karen, this is Bob Bailey. I was going to make a comment here, very similar to the statin measure with the ADA moving towards more individualization of target hemoglobin A1C. So, we may be defaulting back to more profit measure here as well.

Karen Johnson: OK, that was a great point. OK. I think we should go ahead and go to the next couple of measures. They're similar in a way but they do differ a little bit. Luckily, Kyle is the developer (inaudible) here as well. So he – Kyle, tell me if I'm wrong but I think you helped develop these measures as well, right, the 2362 and 2363?

Kyle Campbell: That's correct.

Karen Johnson: So, Dr. Rosenzweig , do you want to go ahead and take 2362?

James Rosenzweig: Sure. 2362 is an inpatient measure and it's an intermediate – it's an immediate clinical outcome measure. It basically – it attempts to identify the average patient – percentage of patient of days of hospitalization of patients with diabetes or who have hyperglycemia and identified the numbers of days − the percentage of the numbers of days that in the hospital that the patient would have high – would be considered to be hyperglycemic.

And the rationale is that there are a number of studies that indicate that there's an association of hyperglycemia with that poor outcomes and also in-patient mortality. And that hyperglycemia, of course, can contribute to making it more difficult to create infection, various other problems in the in-patient setting.

So how I do proceed with the discussion of this? Can we just summarize how the measures actually implemented or …

Karen Johnson: Let's just go through and talk briefly about the evidence. What kind of evidence did the developer provide and – the questions I think we had specifically on this is in some cases were similar to the other measures. Did the guidelines apply? And if not, the other studies, can they be considered a systematic review?

James Rosenzweig: Yes, they identified nine studies. They were not systematic reviews but the individuals studies were large studies that looked at outcomes over a period of time. And they certainly showed in the association of − between hyperglycemia and worse outcomes.

But the question is did they show an association between hyperglycemia as defined in the measure and poor outcomes. And they have a fairly complex algorithm for identifying whether a (day) was considered to have hypoglycemia, sort of six hours – six-hour continuous – six-hour period. They were – blood glucose is greater than 100 – greater than 200 within that specific day.

The other thing is is that the measure is not just applying to patients who have diabetes. It's also has − it applies to patients who don't have diabetes but also have an elevated blood – but have an elevated blood glucose at – discovered at any time during the hospital stay.

So the evidence – so, with respect to this, I don't think the evidence could be considered a formal systematic review. As far as the body of that evidence that were summarized in the document and – what is the – I did not quite understand what a moderate rating was. Could you just explain that again?

Karen Johnson: Sure. Lindsey can you bring evidence algorithm back up? So, basically, while she's bring that back up, if you're fortunate enough to have a measure that's based on a systematic review, that's one of the easiest things that can happen. And then pretty much all the developer needs to do is summarize what the systematic review found in terms of quantity, quality and consistency. But if that doesn't happen, then we have to go a little further and − Lindsey can you slow down a little bit. And going back to the algorithm, since it's not a systematic review, we have to go down to box – our system is a little bit behind. So I know Lindsey is – there we go.

So we're at box seven. So since they did tell us about studies, there's are nine studies, but not a systematic review in grading. So then your question at box eight is, does that include all the studies in the body 11. So the question there is just trying to get to whether pretty much everything was included.

So again, recognizing that we − the measure − there may not be a formal systematic review but still the developers were able to capture all the literature that's out there on that particular measure. And then what does that tell you, you know, in terms of box nine. Do you feel like based on that literature that's provided, did the benefits clearly outweigh any undesirable effect?

James Rosenzweig: Well, based upon – I mean they took the best studies that applied to this, the nine studies that I was aware of. I wasn't aware of any other major studies that identified this in the same way. There are a number of studies that looked at greater blood glucose but not necessarily with days of hyperglycemia as defined in the measure.

Karen Johnson: OK. And really to be honest, what that portion of the algorithm is trying to get to, you don't want, for lack of a better work, cherry-picking of study. So you don't want someone to just pick all the studies and all papers that would support their measure and conveniently forget to tell you about measures that maybe wouldn't support your measure. So, that's what that part is looking at. That's why we want to know about the body of evidence. So it sounds like you're pretty happy that they didn't cherry-pick. Is that a fair …

James Rosenzweig: I don't think − I don't think that they specifically cherry-picked but they − the way that the – but – so studies more identify specific, you know, a relationship between hyperglycemia that may not necessarily be defined in the exact way that the measure was constructed here.

Karen Johnson: OK, OK. And just so we're all clear about that, you know, that does happen because sometimes measures are constructed in certain ways and always be reflected in the literature. And so, you just kind of have to use your best judgment about whether what they found is close enough. And then under validity there is that question about how does the evidence apply to the specs. So, you think about the evidence again when you're thinking about the validity. So, they pretty much do the best they can and then you have to think about whether, you know, if they couldn't find exactly how things are specified in the way they did it, you know, does that impact that validity of the measure itself.

James Rosenzweig: Yes, because, for instance some of the studies identified prevalence of hyperglycemia during a hospital stay but not necessarily the percent of days in which the patient had hyperglycemia.

Karen Johnson: OK

James Rosenzweig: And I give you – that's an example − a way in which the studies were not exactly, you know, it did not exactly fit the actual measures itself, but we're certainly consistent with the overall importance of the, you know, the importance of measuring hyperglycemia.

Karen Johnson: OK, and that's exactly what, yes, definitely we'd like you to think about.

James Rosenzweig: Yes. This particular measure as I said identifies those days in which you have like six hours in which there's continually, you know, high blood glucose. So − and high blood glucose is not that high, they're just over 200. And the question is, you know, obviously, the prevalence of hyperglycemia is important but is the number of days with hyperglycemia really be necessarily the best way to measure this.

Karen Johnson: Right. OK. Let's go ahead and talk about gap, Lindsey, I'm looking at our time. Or maybe as we discussed, were there other pieces maybe not so much gap but some of the other criteria that maybe we should discuss because we don't have time to go through every single one of them.

James Rosenzweig: Well, measures developers clearly identified that there are gaps in care existing, you know, and they are – their own, you know, studies indicated that, you know, the range is somewhere between like 25 percent and 35 percent. I don't have the document exactly in front of me.

But − so there are gaps in care that could be addressed. In addition, there are Hispanic patients specifically think they have much higher blood glucose levels, or seem to have a higher blood glucose levels in general than, you know, hospital.

But, of course, this is an intermediate measure. So it doesn't necessarily mean that − it doesn't necessarily mean that within the specific disparities that they are also – that they correlate necessarily with the, you know, actual outcome measures like morbidity and mortality …

Karen Johnson: OK

James Rosenzweig: … for each individual group. In other words, it doesn't – the evidence doesn't necessarily show that Hispanics per se with, you know, with hyperglycemia have more mortality or more morbidity. Just the evidence indicates that Hispanic patients have higher degree of …

Karen Johnson: Hyperglycemia.

James Rosenzweig: … of hyperglycemia, yes.

Karen Johnson: OK. Let's go ahead and skip ahead because priority, it's almost (a given), I think. So, let's talk about specs more briefly. Did you have any questions or concerns about the specifications?

James Rosenzweig: Specifically – the one concern I had was the way they defined in the denominator patients that just have one blood glucose is greater than 200 being part of the denominator. I mean, there can be abnormal blood glucoses of greater than 200 in a lot of patients due to a variety of settings that wouldn't necessarily apply. So you're probably including a large percentage of patients who don't have diabetes but incidentally had a blood glucose greater than 200 which would automatically lower your percentage of patients, meaning full patients with hypoglycemia.

Female: Yes, I probably like to see it as another confirming blood glucose just because of error, human error of taking that one blood glucose especially …

James Rosenzweig: Yes.

Female: … in the hospitals, especially with – in a teaching hospital with medical assistants and unfamiliarity with a hospital (neither) versus what's used out in the real world.

James Rosenzweig: I would agree, yes. So that was the – that was one specification issue that I think it would need to be addressed. The other one is just perhaps some standard – whether or not there might differences in actual recording of the, you know, the measures themselves. You need to – I mean in order – to be able – you need to have assistants that will make sure that they are recording accumulate all of the blood – finger-stick blood glucoses that are done in order to compare different hospitals with their – for their outcomes.

And, of course, if there are patients that don't have blood glucoses measured very much, they may have a lower percentage of – they may have a lower percentage of days in which blood glucoses or hyperglycemic partially because they missed them.

Female: OK.

James Rosenzweig: And the other thing is – the other thing is that a number of studies that have applied to this did segregate out in − ICU patients compared with the total patient population. And the issue was of, course, that within the ICU, you know, blood glucose could be measured extremely frequently.

Female: Right.

James Rosenzweig: And there could be periods that would – so you would be able to identify hyperglycemia much greater during those periods of time. And it might be – at least so you know some of the studies might indicate that patients in the ICU setting might be more affected by – by extended periods of hyperglycemia. So look, I don't have an answers to whether or not they should actually − this should be reported with respect to, you know, percentages in the ICU setting versus others. But, it's something – I don't know – certainly it came to my mind and I don't know I have to think about this just a little bit more to discuss, you know, at the actual meeting.

Karen Johnson: This is Karen again from NQF. And actually that was one question that I did have in terms with specifications. And I wonder if Kyle, could you spend just a minute or so? It's just this came up. You indicate the measure is risks stratified and you have it by ICU, non-ICU med versus surg. And I think there's a third one that I've forgotten.

But it's clear, because you almost showed data that say maybe you don't need to stratify in those ways. And in general, we talk about stratification with verifications, it wouldn't be a three separate stratifications. So I got confused and I wasn't sure if you were saying it didn't need to be stratified or perhaps it should be reported in these strata separately. So could you maybe talk about that just a little bit?

Kyle Campbell: Sure, I'm going to briefly respond and then I'm going to defer the question to Almut Winterstein, pharmacoepidemiologist at the University of Florida and help us to construct the constructive measure.

Just in terms of the stratification, we we're recommending or asserting that the measure being stratified by ICU and non-ICU for the reasons that were mentioned and also because the level of evidence in the different unit is different. In addition, another factor of impact in terms of risk was steroid. And so, we did stratify by steroid, basically, accumulated daily dose to prednisone equivalent because we found much higher rate of hyperglycemia and those (inaudible) high dose steroids.

And with that, I think to best speak to this, I would defer to Almut and hopefully by now she has some – has been and able to get off-mute.

Almut Winterstein: Yes, actually did get off, I think, can everybody hear me?

Female: Yes.

Almut Winterstein: OK, wonderful. And I sort of didn't hear what Kyle said because I was speaking to the operator to unmute me, so please forgive if I'm redundant.

Yes, so there's three stratification rubrics that we are proposing. One is medical versus surgical patients, one is ICU versus non-ICU environment and one is steroid versus non-steroid exposure. And also we have – we have had a lot of discussions in our expert panel meetings to see how best to deal with this.

With respect to medical versus surgical patients, the level of evidence is a little bit different for those who – for surgical patients that is much larger numbers of studies, and probably also more rigorously conducted studies and large to set guidelines for blood pressure – sorry blood sugar controls should lie.

And the link of surgical complications as they are – there are some medical patients, the evidence agrees that blood glucose should be controlled, but the link to complications is probably a little bit weaker in comparison. But I have one reason why we thought those two should be stratified, to allow essentially the integration of evidence and interpreting the measure.

As for ICU versus non-ICU, this is actually quite interesting to see if you look at measure results. On one hand, patients in particular for surgical patients, they'll have greater needs of blood glucose control so they're not even talking about diabetes patients, they're talking about patients who are post-surgical, have high stress levels might have been exposed to high doses of steroids and just have, you know, under chronologic response and in need to a glucose control, so that means it's a very difficult to control population on one hand.

On the other hand, the ICU environments have the ability to control glucose levels much better through insulin drips, so continuous insulin infusion just because the nursing staff to patient ratio was much higher and the ability to follow glucose levels tighter – is much higher.

So that said, we know that in some hospitals that the ICU have actually had better measure rates than the non-ICU. And in some other hospitals, it was the other way around which suggest to us it really depends on how much effort has been put into place in the ICU to control glucose in a more aggressive fashion or not.

So that's what the main reason why we saw it does makes sense to break this down because it is not consistently but we can conclude that an ICU glucose control is worse or better.

And then finally, for the steroids, steroids have a very, very tight correlation as glucose controller as many I'm sure on the call can appreciate, and that relates in particular to a very high doses of steroids, to give you a little bit appreciation for this. And we looked at this in our formative testing, patients on low dose steroid, so less than 40, 50 milligrams prednisone equivalent had an average measure, or rate of about 15 percent, 20 percent similar to the overall population that you see here.

Then as you go to up 500 milligrams, that percentage goes up in a 40 percent range and if you go even higher to 1,000 milligrams of prednisolone, you were up to 60 percent of all measure days being uncontrolled.

So when we talk to the expert panel, I was looking at the literature whether glucose can be easily manage or basically whether blood glucose can be managed in patients on steroid. What we noted was that, well, in general, our experts agreed that high glucose levels can be managed by increasing insulin doses or any kind of anti-hyperglycemic medication for that matter.

With steroids, sometimes the individual patient's response to steroid is unpredictable, which basically means there was a delay in being able to titrate insulin needs up just because it is sometimes hard to predict what's going to happen.

On the other hand, many of our experts agreed that once there was a little bit of experience gathered with this, it's usually possible to control subsequent days. So essentially, the conclusion was that it might be difficult, more difficult to manage patients on high dose steroids but not completely impossible that therefore the measures would be stratified to essentially show this and allow comparison of institutions.

Karen Johnson: OK. So, I'm sorry, we're so close on time.

James Rosenzweig: Yes. I have to honest. I missed the stratification discussion. I can't even find it now. Where – actually if the staff could send specific information related to stratification to me after the meeting, I'd appreciate it. Thanks. But, I mean, this is very helpful.

Karen Johnson: Yes. I think, and maybe we can talk offline just a little bit. So Kyle and company − so it sounds like you're saying, even though your results are a little bit iffy on a couple of them, you do want to suggest that everybody computing this measure would stratify by the three separate things. Is that a yes or no?

Kyle Campbell: Yes. For certain, we're recommending ICU and steroid. And I think medical and surgical, we we're hoping to see, you know, what the thoughts of the steering committee were.

Karen Johnson: OK.

James Rosenzweig: And the stratification would apply if the patient just happened to be in the ICU during – once during they, let's say, during the hospital stay, if it would apply to the entire amount of time or is it just to the ICU stay time?

Almut Winterstein: It's just the ICU stay.

Kyle Campbell: Yes. (Inaudible).

Female: And what do you do to the ICU – post-surgical ICU patients on high dose steroids? Where do they stratify, too?

Almut Winterstein: Yes. I mean, that's always an issue, you know, the most stratification sectors, they are the more complicated as it becomes. The flip side to this would be to build a complete risk adjustment model …

Female: Yes.

Almut Winterstein: … that, you know, looking at the risk adjusted model for a surrogate outcome where there is fairly strong agreements that this outcome is manageable including insulin need. If we we're building a risk adjusted model that predicts hypoglycemic days, what would go in there is whether I'm post-surgical and whether I am on steroid and whether I'm older and whether I have diabetes and whether my age or A1c was already uncontrolled in the out-patient environment, and a variety of other intrinsic factors, which in reality can be completely controlled.

And, you know, this is where the risk adjustment model, I think, becomes inferior because if we we're looking at the validity or a risk adjustment model measured by, you know, basically (inaudible) statistics to say, you know, this is a fairly good prediction of what's going to happen. We would model essentially some of the quality issues (inaudible). Because many of the patient intrinsic factors that results in high glucose can be managed.

And I think this is quite unique to, you know, the management of glucose, you know, similar would be probably the management of INR or, you know, managing − any kind of lab for that matter.

Karen Johnson: So this is Karen, just a couple of things from the NQF perspective. When we talk about risk adjustments, we recognize a couple of different ways to do risk adjustment. One is through statistical modeling like you just talked about. And another is through stratification. That's just another way that you can risk-adjust.

So, I think we need to make sure that we're understanding that you're suggesting risk adjustment through stratification. And then at some point maybe between now and the in-person meeting, we probably need to get straight some of these things. Generally, when you talk about risk adjustment, you're doing – you're talking about things as start of care, but you know, the ICU versus not – med-surg versus not may not be obvious at the start of care. So that's a question that you will probably have to address at some point.

Female: Yes. In my experience, a high blood glucose sends you to the ICU so that they can have better control over it post-surgically.

Karen Johnson: Yes.

Female: Whether you have diabetes or not, so.

Karen Johnson: Yes. So these are certainly questions that I think we have to …

James Rosenzweig: And there's specific surgical procedures, you know, especially cardiac surgical procedures that have – inherently will have hyperglycemia more often.

Female: Yes.

James Rosenzweig: I would be – yes. I don't know – I have to be honest, I mean, I think the ideas of certainly splitting them up into individual baskets like that is a good idea. The question is, I don't know whether you could really develop − you'd have to really show very carefully the − develop a system of risk – not risk stratification but of a system of being able to risk-adjust appropriately. You have enough data for that purpose.

Almut Winterstein: That's the beauty of this particular measure. There are so many people in the denominator and so many days that are hyperglycemic that it is actually possible to slice in quite a number of pieces. The question is, you know, how confusing it becomes when this is done.

James Rosenzweig: Yes, I just – I would – even though there's so many patients that could be in this, the question is can you really come up with a composite score or should you leave the individual scores of the patients and the various groups separate?

Anne Leddy: Karen, this is Anne Leddy speaking. I haven't had a chance to discuss my topic. And unfortunately, we have – my husband has an international conference call for the ASBMR in a few minutes. So if I just like to break in and raise a few points about the last topic which is 2363 glycemic control …

Karen Johnson: Great, yes, thank you.

Anne Leddy: Everybody can easily see what measure is and what the specifications are. My concern after listening to the prior discussions is, is the body of clinical evidence adequate. I do believe the studies used and they aren't very many were well-discussed. I think this is an important measure. The incidence is quite low but it does represent a huge problem related to errors in medication in a hospital setting. So it is a very important measure. And I think it's important that if they link to the hyperglycemia measures.

Karen Johnson: Thank you, Anne.

Anne Leddy: I would love to talk more and get your help on this validity, reliability which is my weak area. But that will have to wait until next week.

Karen Johnson: OK.

Male: I agree with the ideas linking at myself because, you know, the data accumulation would be very similar for both groups.

Anne Leddy: And of course the treatments – aggressive treatment to hyperglycemia is going to result in more hypoglycemia.

Female: Yes, yes.

Anne Leddy: So (inaudible) for that. And then just one last comment about the indicator of less than 40 versus less than 70, if we choose less than 70, there going to be many, many more incidence but perhaps the caregivers will then review full close records more systematically and avoid severe hypoglycemia. But again all of these needs to be discussed in greater detail. So I'm going to sign off and look forward to seeing you all next week.

Karen Johnson: Great. Thank you.

Anne Leddy: OK, bye.

Karen Johnson: Bye-bye. Does any of the other committee members have any burning questions or concerns about either the hyper or the hypoglycemia measure that we haven't already discussed? I think for the hyper, we haven't really talked about the cut point being greater than 200. So that might be something that we have to discuss in the in-person meeting.

Robert Bailey: And this is Bob Bailey. I guess the other consideration here is one measure, the hypoglycemia one includes metformin and the hyperglycemia does not – I understand the reasons why, but there's not harm to the (patient here).

Karen Johnson: OK.

Male: That's a good point. My overall concern is you have a 200 – a single blood glucose of 200 is just − it happens too commonly in the in-patient that might not really be appropriate for the measure.

Karen Johnson: OK. We'll be sure to get that point down in our summary.

OK. In just a minute we'll then open this call up to public − any public comments that we may have. But I will encourage the committee if you have questions as you're going through these or other measures before next week and you just have questions about the algorithms or any of their process rules, or anything like that, please just send us an e-mail and we'll feedback to you as soon we can.

So Lindsey, I'll turn it over to you open it up.

Lindsey Tighe: Yes. Operator, if you could open the lines for anyone with a comment.

Operator: At this time, if you would like to ask a question, press star one on you telephone keypad. We'll pause for just a moment to compile the Q&A roster.

And there are no questions at this time.

Lindsey Tighe: All right. Well, thank you very much committee members and also measure developers for joining us on today's call. We look forward to seeing all of you in person next Wednesday and Thursday in Washington D.C. If you have any questions about the travel logistics, please let us know. And we do plan to send out a final agenda later today.

So we'll get that out to both the developers and committee members so you'll see what order you're presenting in and certainly we just ask that all of you workgroup members be prepared to walk the entire committee through a similar conversation that we've had today.

As Karen mentioned, we will be providing summary of this call to you in advance of the meeting. So thanks again, and please don't hesitate to contact us if you have any questions between now and next week.

Female: Thank you.

Male: Thank you.

Female: Thanks.

Female: Bye.

Female: Bye.

Operator: Thank you. This concludes today's conference call. You may now disconnect.

END