NATIONAL QUALITY FORUM

Moderator: Infectious Disease
March 1, 2017
1:00 p.m. ET

OPERATOR: This is Conference #: 92638622.

Operator: Welcome everyone. The webcast is about to begin. Please note today's call is being recorded. Please standby.

Christy Skipper: Good morning – good afternoon everyone. Welcome to the second work group call for the Infectious Disease Standing Committee. My name is Christy Skipper, Project Manager and I want to turn it over to my team to introduce themselves.

Mauricio Menendez: Hi everyone, Mauricio Menendez, Project Analyst.

Melissa Marinelarena: Good afternoon everyone, this is Melissa Marinelarena, Senior Director on the Infectious Disease Project. I'd like to welcome everyone this afternoon. We can discuss the two Sepsis measures that we're going to be reviewing during the project. And again, thank for joining us. I will turn this back over to Christy.

Christy Skipper: Thank you, everyone. So, as Melissa said, the purpose of the call is to run through or review the Sepsis measures submitted to this project. And the format for today since we – some of you received the preliminary analysis from measure 3215 just a couple of days ago. We want to start up by asking our developers to introduce their individual measures and then turning it over to the committee to ask any questions of the developers.

And before we actually get started with that, I'm going to turn it over to Mauricio to call roll. So when you here your name, just please say here.

(Off-Mic)

Mauricio Menendez: Woody Eisenberg?

Woody Eisenberg: Here.

Mauricio Menendez: Adam Thompson?

Adam Thompson: Here.

Mauricio Menendez: Emily Aaronson? Amesh Adalja?

Amesh Adalja: Here.

Mauricio Menendez: Esther Babady? Nanette Benbow? Kathleen Brady? Laura Evans? Piero Garzaro?

Piero Garzaro: Here.

Mauricio Menendez: Donald Goldmann? Jeffrey Hart?

Jeffrey Hart: Here.

Mauricio Menendez: Michael Lane?

Michael Lane: Here.

Mauricio Menendez: Jeffrey Lewis? Melinda Neuhauser? Rocco Orlando?

Rocco Orlando: Here.

Mauricio Menendez: Jamie Roney?

Jamie Roney: Here.

Mauricio Menendez: Pranavi Sreeramoju? OK. Thank you.

Christy Skipper: All right. And then also I like to acknowledge any developers from either the New York State Department of Health or the Henry Ford Foundation, or other, if you could please call out your name if you are on the phone?

Ed Septimus: Ed Septimus is on the phone.

Christy Skipper: Yes. And thank you Ed for joining us. The co-chair of the Patient Safety Committee.

(Crosstalk)

Christy Skipper: I'm sorry. Could you do one at a time, please?

Emanuel Rivers: Emanuel Rivers from Henry Ford Hospital.

Christy Skipper: Hi, Manny.

(Kim Servollo): (Kim Servollo) from Henry Ford Macomb Hospital.

(Jennifer Wisnoski): (Jennifer Wisnoski) from Henry Ford West Bloomfield Hospital.

Foster Gesten: Hi, Foster Gesten from the New York State Department of Health.

Christy Skipper: All right. So thank you all for joining us today. And I just want to remind you that you should see both logged in to the webinar and dialed in on the phone in order to fully participate.

So we're going to start this afternoon with Measure 3215, Adult Inpatient Risk Adjustment Sepsis Mortality. This is a new measure submitted by New York State Department of Health. So, we'd like to turn it over to the developer of this measure to give us a brief introduction to the measure.

Foster Gesten: Sure. This is Foster Gesten. You guys can you hear me OK?

Male: Yes, we do Foster.

Foster Gesten: Terrific. I have some colleagues. I don't think they introduced themselves so I'm hoping who are on the phone who helped us in the development of this measure. They include colleagues at IPRO, at Brown, at Ohio State

University. So they may jump in and help with answering of questions or provide some overview.

But very briefly, first, you know, thanks for the opportunity to submit this and review the measure. This measure is, just in terms of background, it's part of one component to an overall initiative that we have in New York over the past few years to try to improve outcomes for patients with sepsis. It includes requirements on hospitals to have protocols for early recognition and treatment of patients with severe sepsis and septic shock, as well as provide data to the department to evaluate the implementation of those protocols as well as outcomes for patients.

And obviously, the outcomes part, the risk adjusted mortality part for adults that we're here to talk about. This measure was developed as part and parcel of that initiative which is both a public reporting as well a quality improvement initiative in the state.

We focused on adult patients and make use of the data that was collected an ongoing basis through a portal that IPRO helped us develop that gathers the patient specific information that enabled us to working with one year's worth of data developed a model for risk adjusted mortality.

We had the ability to do some testing, validity testing which I think is in the document. And if (Diane) or Gary are on the phone, you know, they're available to answer some of the questions of that – the model itself and the testing. And (Kathy) can answer. We can answer some questions about the auditing of the data, reliability of the data as well. I can see that came up as an issue in terms of their worksheet.

The hospitals have been getting feedback on their performance since the beginning of the initiative including the ability to look at the raw mortality compared to state-wide. Only recently, in the past months, hospital had been able to look at their risk adjusted mortality rates as well. And our plan is for public release actually some time this month, potentially in the next week or so of the information on our website and it's part of the event.

So, I don't know if you're looking for more or put information Melissa, in terms of overview, but why don't I stop here and see if that's enough so we can get on.

Melissa Marinelarena: That's great. Thank you, Foster. We can start the conversation. The committee can start with questions or if you prefer, I can provide some clarification over the rating on the validity section on the inter – on the data element validity section, for those of you who have had the chance to look at it.

Woody Eisenberg: Melissa, this is Woody. I think it would be helpful if you could read back to us some of the questions and comments that we submitted. I for one don't have my own record of that but it sounds like you do and as Foster's been able to see it already.

Melissa Marinelarena: Sure. And it's up on the screen. So, this is an outcome measure. So the – for evidence it's a little bit in process measures. And for those of you that were on the call yesterday, we had one outcome measure. So for an outcome measure, we just asked if the committee agrees that the developers provided a rationale that links the outcome and processes that either a provider or a hospital, or a process that can be changes something that the provider has influence on the outcome, and that's a yes or a no.

And so, for those of you that are on the platform, some of the committee comments included for evidence it says, this measure just requires reporting of mortality which is a PRO but isn't tied to a specific process or intervention. I agree that at least one hospital process identified by the developer impact inpatient sepsis mortality and yes, strong and well substantial relationship.

And again, we just sent this out. It was early Monday morning. So, we understand that you haven't had chance to look at it yet so that was why if anybody had any questions or we just wanted – you wanted to have a discussion on it, that was the purpose of this call today.

So if anybody has any questions about evidence, and again, it's just the rationale and then this is a past (inaudible). If not, we can talk about the

performance data that they provided based on the one year data that they provided, if you can scroll down.

Emily Aaronson: Sorry I just joined this call a little late. To clarify, are we right now on the outcome measure on the 500 or on the other?

Melissa Marinelarena: On the outcome measure.

Emily Aaronson: OK.

Melissa Marinelarena: Is this is Laura?

Emily Aaronson: This is Emily, sorry.

Melissa Marinelarena: Hi, Emily.

Emily Aaronson: Hi.

Laura Evans: Hey, this is Laura. I joined a couple of minutes late too. Sorry about that.

Melissa Marinelarena: Hi, Laura, welcome. And Pranavi, are you on the phone as well?

Pranavi Sreeramoju: Yes, I am. I wasn't able to complete the survey but I do have some question.

Melissa Marinelarena: That's fine, we understand. Probably all of you have the chance to complete it before the meeting, so that we can share your questions to the developer at the time and with the rest of the committee. So this was – we're able to provide some database on the one year that Foster mentioned. So there are some risk adjusted mortality rates that are up on the screen. I don't know if anybody has any questions for Foster and his team regarding this.

Laura Evans: I don't any, this is Laura. I don't have a question about this. I would, you know, sort of echo the preliminary comments, you know, the evidence supporting evidence that can move risk adjusted mortality as a marker of sepsis quality, I think is quite robust.

Melissa Marinelarena: Thank you. If you scroll down.

We're also able to provide some data based race, ethnicity, gender, age, and insurance and payer. So they were able to break that down. And if you scroll down, performance gaps. Some comments included at the data demonstrate variation and less than optimal probability as inpatient sepsis mortality rate. There is a gap in care that warrants a national performance measure. Subgroups are identified. Disparities are present but limited. And this was not a composite measure. These are responses to our standard questions that were in the survey.

Are there any questions, comments?

Pranavi Sreeramoju: Yes. So this is Pranavi. So, you know, I read – I fast read most of the material that was shared with us and it's definitely my first time doing this type of exercise. So, pardon me if I sound very ignorant but I do have a few questions.

So, you know, I understand the measure and I'm intimately familiar with the NQF's CLABSI/ CAUTI and sepsis bundle adherence and necessary measures. So this one is related to the sepsis bundle adherence. So, the base for this measure, the basis for this is the New York data. So, how this measure going to be used? What's the proposed use for this measure? Is it going to be the CMS code measure data that all hospitals are reporting for sepsis mortality? Is it something else? I think that will help sort of channel the talk processes, so that's one. And I don't know if should share all my questions with you so that you can sort of organize the conversation. I don't know. Do you want me to share all the other questions I have on my mind?

Melissa Marinelarena: I can respond and give you the NQF perspective and then I let Foster respond and his team of how they are using and how they plan to use it.

Pranavi Sreeramoju: OK.

Melissa Marinelarena: As far as how an NQF endorsed measure is used …

Pranavi Sreeramoju: Yes.

Melissa Marinelarena: … for CMS, we don't decide that here. That's, you know, that is the separate process.

Pranavi Sreeramoju: OK.

Melissa Marinelarena: We do asked our developers to provide if the measure is in current use and what kind of plan they have for it in the future, as far as quality improvement and accountability and public reporting programs.

Pranavi Sreeramoju: Right.

Melissa Marinelarena: But we don't – this is separate from the CMS measure.

Pranavi Sreeramoju: I see.

Melissa Marinelarena: So that's – that is a separate process. But now I will turn it over to Foster and his team and he can give you more details about how they're using it, how they plan to use it.

Pranavi Sreeramoju: OK.

Foster Gesten: Yes, sure. You know, reasonable question understand the context. I said in my opening statement that the purpose of this measure for us in New York is partly partial of the series of measures that are part of our sepsis improvement initiative.

Pranavi Sreeramoju: Right.

Foster Gesten: So, specifically, we're using this as a way for hospitals to be able to get a better sense of where they are relative to the care they're providing in the hospital.

Pranavi Sreeramoju: Right.

Foster Gesten: They use this in conjunction potentially with other measures they may have that may include data if you're getting it on SEP-1 but we also are tracking other measures that look at three-hour and six-hour performance as well as the use of a protocol.

Pranavi Sreeramoju: Right.

Foster Gesten: So, it's a combination of, you know, purpose for internal quality improvement for the hospitals as well as public reporting and transparency …

Pranavi Sreeramoju: OK.

Foster Gesten: … which is work the initiatives that we have. I can't, you know, as Melissa said, I can't really speak for how it might be used by others including CMS.

Pranavi Sreeramoju: OK. So, Foster, has it been shown that this metric responds to intervention? I mean, the main thing in quality improvement is responsiveness of a certain metric to interventions. Has that been shown based on the New York data or anywhere else?

Foster Gesten: I'll answer in one way and maybe others can, Mitchell if you're on, might answer a slightly different angle. What I would say is we have evidence although we wasn't presented in the context of the measure to suggest that changes or improvement in use of protocols and/or the timeliness of those interventions does in fact have a significant relationship to mortality.

So, if your question is, you know, is there evidence that there are interventions that can change or impact mortality, I think the answer to that is yes. And we see it both in, you know, the literature and another initiative as well as in our own. But Laura or Mitchell might have some thoughts about that as well. Did answer your question?

(Crosstalk)

Pranavi Sreeramoju: So, I'll just add more context to my question. I definitely understand that the measures responding intervention in general but my question is related to this specific metric because risk adjusted mortality. So to give you an example, I'm a health care epidemiologist to the large safety net hospital and we have a sepsis mortality reduction initiative. And our sepsis bundle adherence for all patient, sepsis, severe sepsis and septic shock went from 14 percent to 35 percent over four years. And our sepsis mortality grows

mortality, numb crude in hospital mortality went down from 9.8 percent to 2.3 percent in the same period.

So, I'm fully onboard in terms of yes to crude mortality or, I mean, outcomes do respond to interventions, but my questions related to this specific metric. Has this been measured against interventions that were undertaken?

Mitchell Levy: So, Foster, if you want, this is Mitchell Levy …

Pranavi Sreeramoju: Yes.

Mitchell Levy: … I'm working with the New York State Department of Health. And I think I'm about to reiterate what Foster said that …

Pranavi Sreeramoju: Yes.

Mitchell Levy: So in the database that we have which is a 100,000 patients …

Pranavi Sreeramoju: Right.

Mitchell Levy: … over the two – over the two-year period …

Pranavi Sreeramoju: Yes.

Mitchell Levy: … we were able to demonstrate that as compliance went up …

Pranavi Sreeramoju: OK.

Mitchell Levy: … over two years and we also looked at core trials of compliance …

Pranavi Sreeramoju: Right.

Mitchell Levy: … and we're able to show that over time in the core trials of compliance …

Pranavi Sreeramoju: Yes.

Mitchell Levy: … there is a statistically significant association with decrease risk adjusted mortality using this metric, this measure.

Pranavi Sreeramoju: OK.

Mitchell Levy: So, I think the answer is yes to your question.

Pranavi Sreeramoju: That's helpful. That's helpful. And did you have – I think that there are 16 measures in the prediction model. Do you have (fee) statistic in terms of what percent of the outcome improvement is attributed to these 16 measures? And then on the same team, my other question is do you have organizational characteristics in the prediction model? I see more individual-based variable in this.

I don't know that I saw organizational characteristics and – especially someone who works predominantly in a safety net hospital. It's not just individual young patients – a majority of our sepsis patients are actually very young and that's why our overall mortality in the single digit, but at the same time that are all these organizational characteristics that I feel do go into the outcomes and I – do you have any data on that?

Foster Gesten: So we have – we'll go into your first point. We have …

Pranavi Sreeramoju: Yes.

Foster Gesten: We have information in coefficients and …

Pranavi Sreeramoju: OK.

Foster Gesten: … you know, impact for each of the variables which I think is your question is …

Pranavi Sreeramoju: Right.

Foster Gesten: … how much they contribute to the model.

Pranavi Sreeramoju: That's right.

Foster Gesten: So that information is available. The second question about organizational characteristics, we – for purposes of evaluating the hospital performance, we have made a decision and it tracks to all the other public reporting that we do

whether it's risk adjusted mortality or otherwise, that we do not adjust for organizational characteristics …

Pranavi Sreeramoju: OK.

Foster Gesten: … which is not to say that's not a research interest or that we don't have data to look at …

Pranavi Sreeramoju: Sure.

Foster Gesten: … a hospital characteristics and so on from a research agenda. But for purposes of the model and the context of evaluating …

Pranavi Sreeramoju: Yes.

Foster Gesten: … absolute performance, we tried hard to not include interventions in the model or – and did not – and chose not deliberately to include hospital characteristics but you're absolutely right, that they can make a difference and we're interested in understanding what does how they've influence the outcomes.

Pranavi Sreeramoju: OK. I'm interested in knowing – learning what others feel about that. You said there are data on organizational characteristics but you left them out, am I understanding you correctly or they've not been explored fully enough to go into this measure which one?

Foster Gesten: We deliberately chose not to include them in the model.

Pranavi Sreeramoju: OK.

Foster Gesten: We have certain characteristics for hospitals. You know, their type, their, you know, urban/rural, teaching and non-teaching and so on. And it's part of a research agenda we're looking and trying to understand what are the hospital characteristics as well as patient characteristics that maybe associated with better outcomes or better performance.

But for purposes of the mode looking at risk adjusted mortality and I think this is not unusual in models like this.

Pranavi Sreeramoju: OK.

Foster Gesten: Hospital characteristics were not taken into consideration in the development of the model.

Pranavi Sreeramoju: OK. Are there data that you can share with me or in our group do you think?

Laura Evans: This is Laura. Can I maybe add a couple of comments as somebody has some experiences as a end user of this model, being hospital in New York State.

I think this a really interesting discussion. I would just sort of add from my perspective as somebody who has led our internal sepsis improvement efforts out of New York State hospitals. So I have sort of experience with submitting data to this model as well as receiving reports from this model on our hospital performance.

I think there are several things that are pertinent to this discussion and similar I think to your setting. I practice in a safety net large urban, public hospital setting in New York. And I think one of the things that's been interesting about this journey for us internally has been that the risk – the ability to discern risk adjusted mortality has been actually very critical to us for internal improvement efforts because I don't know but other safety net hospitals but we often don't perform very well in quality measures. And often you'll get sort of this argument from staffs at the hospital that's because our patients are different because they're younger or sicker or less access to care and all of this – which are undoubtedly true.

So being able to pull out risk adjusted mortality that's demonstrably, you know, sort of benchmarks our performance relative to other hospitals New York State, that's actually been very important for our internal improvement efforts.

And so, we locally have actually found this risk adjustment model to be very useful in terms of understanding our performance and seeking to improve it.

So I just thought I would sort of throw that in there, somebody who has some experience of actually using the model.

Pranavi Sreeramoju: So it's actually helpful to hear your perspective and your experience. So, I'm hearing that – you've used this for internal quality improvement. What about external generalizability because in comparisons, I mean to me, a lot of times risk adjusted mortality or risk adjustment outcomes are more helpful for external comparisons, and not to say that they're not helping for internal comparisons but they're quite helpful for external comparisons comparing with other organizations.

And like you just noted, for safety net hospitals the comparison with other hospitals runs into issues of resources, issues of demand like busy emergency departments and things like that. And it's not necessarily because the patients are younger or sicker but it's – I mean younger patients actually have less more mortality. You have less sick patients coming to the E.R. So, you – they should see less mortality. But the hospital is so busy and so chaotic. There are 60 patients waiting for admission so that contributes to the less care.

So, you run into a very complicated situation and that's why, I mean, I would love to see a risk adjustment variable based on organizational characteristics for external comparisons. But internally, yes, it makes sense.

Male: The other thing is there is …

Pranavi Sreeramoju: Yes.

Male: … a people outside that are having difficulty …

Pranavi Sreeramoju: Yes.

Male: … with the audio.

Pranavi Sreeramoju: OK. I can hear perfectly well.

Christy Skipper: So thank you. I just want to make an announcement that you should be logged in to both the webinar on your computer and dialed in on the telephone and I'll also send the message into the chat room with the dial in number.

Mitchell Levy: It's Mitchell Levy again if I can jump in. I think that the whole point of not including the organizational dynamics is creating an independent risk adjustment model but then allows like institutions to compare their risk adjusted mortality and then look at where they stand would like institutions, otherwise your models becomes too complicated because of the hospital characteristics and it's difficult to interpret.

Rocco Orlando: This is Rocco Orlando. I'll start with a comment and then a question about the risk adjustment. So the comment is – so I – in part of five hospital health care system in Connecticut where we've actually been using the very similar model for the last two years. And so the answer is yes, it does drive performance improvement associated and we haven't publish it yet but we do see clear links between adherence to the sepsis bundle and gaps and care and how we're performing and have driven improvements. So, we can – certainly – our experience can support to the approach.

The question is in terms of the risk adjustment as I look at through, you know, the patient sites, socioeconomic determinants of health payer status, all of those things. The one thing that is excluded from that which makes sense was inter-hospital transfers which, you know, clearly is an organizational characteristic of sorts in terms of some places get a lot and some are exporters of patients and not receiving them. So that makes sense.

So my question for the, you know, for the measure development team is, were there other measures that approach that level of significance? Because clearly when you look at the data that was the – that certainly stood out as an important predictor of outcome.

Foster Gesten: Just so I – make sure I understand the question. Are there things that were close to significance related to organizational features?

Rocco Orlando: No. For patient characteristics that were confounders, I mean again, from getting the data late, you know, I scanned it, and it didn't look like they were. I was wondering if there are other things that begin to get close to that level.

Foster Gesten: Yes. Mitch, I don't know if you recall top of your head. And Gary Phillips who from Ohio state I think is having some trouble getting an open line. He might be able to answer.

But, I think there were other variables that we looked at. The ones that we chose or ones that reached significance and, you know, that made sense, and again contributed to the model. There were some others that may have approach significance that we drop. I'm not thinking – I'm not able to – I'm not sure what they've – I can't sight what they might have been off the top of my head.

But the issue that you described in terms of the transfer issue was clearly one that's – was very much on the mind of hospitals as we talk to them about the development both the initiative of the model with …

Gary Phillips: Hello?

Foster Gesten: … a clear suspicion that that's contributed. Hi, is that you Gary?

Gary Phillips: They finally give me an open line after six or seven tries.

Foster Gesten: Great. Gary if you hear the last – the question for the last question about other variables that approach significance that we're not in the model. Do you – I don't know if you have any answer or can recall off the top of your head other variables that were not in the …

Gary Phillips: Yes. There were lots of variables that approach significance but some of them, you know, we've removed because we felt they were over fitting the model. They've had correlation with other …

(Off-Mic)

Gary Phillips: So we try to put in ones that were highly significant and also make clinical sense, at the same time and aborting what we call variance inflation where two particular variables are highly correlated.

Jamie Roney: This is Jamie Roney and I just have a comment. Dr. Levy, I appreciate your clarification on why you left out the institution characteristics for comparison in compared to the analysis purposes. So thank you for clarifying that. That's the only comment I have. Thank you, sir.

Mitchell Levy: Thank you.

Gary Phillips: Hello?

(Off-Mic)

Pranavi Sreeramoju: Yes? Sorry, maybe no one else is talking. So I have another question. This is Pranavi. So, have you noticed of the – the question is for the measure development team. Have you noticed that the responsiveness to intervention is different in different organizations and what are the attributes of organizations that responded really well to the interventions as supposed to those that did not? I mean can you shed some light on it?

I know the organizational characteristics been going to this measure but I'm just trying to understand the future of this measure.

Foster Gesten: I think – this is Foster. The best answer I can give you. It's a great question and one that we're anxious to understand as well is too early to really be able to tell.

In the intervention – you know, even defining what's the intervention was 2013 when they – when the regulation started, was it in 2014 when they measure it, when that data collection started, is the intervention really starts once public release happen this month. I mean, we're able to look at a trajectory of change of over time.

Pranavi Sreeramoju: Yes.

Foster Gesten: We have not analyzed that trajectory of change by hospital characteristics to be able to say who has improved your change more or less and others for example which I …

Pranavi Sreeramoju: Got it.

(Crosstalk)

Foster Gesten: … trying answer your question. So it's a really great question and it's one that I think we have lot of interest.

Pranavi Sreeramoju: Yes.

Foster Gesten: Not only, you know, from the research point of view but from trying to understand how we can facilitate improvement, you know, state-wide and raise the bar across different kinds of facility.

Pranavi Sreeramoju: Sure. And I'll explain a little more about – more on where I'm coming from. You know, I mean most of you are familiar with the health associated infection data from NHSN and, you know, the risk adjustment, I mean there are some organizational characteristics in there.

And we have data from thousands of hospitals nationwide and their use for pay-for-performance. And there is a general sense among hospitals and the hospital leaders and also frontline clinicians although the sepsis mortality and sepsis bundle adherence measures right now are pay-for-reporting.

So there is an incentive for reporting in data collection and reporting but most measures that – actually all measures that are now pay-for-performance started as pay-for-reporting some time ago, a decade ago or a few years ago.

So, there is a general sense that the sepsis mortality measures will go in the direction of pay-for-performance in the future, in a few years from now.

And it would be very important for us to know the impact of not just patient level characteristics but also the organizational characteristics. I mean especially safety net hospitals are already under the weather for – from a value-based purchasing, for all the HAI and other performance measures right now. So, this might become an added measure that they get penalized for. So that sort of overarching concern that I wanted to share with you and that might help you understand where I'm coming from.

(Crosstalk)

Foster Gesten: Go ahead.

Male: Gary can you clarify were there no hospital level characteristics that wound up being significant and incorporated in the model? Gary?

Foster Gesten: Gary, you may have (it) on mute. Go ahead.

Gary Phillips: Yes, I was just going to say, we did that (for any), I believe …

Male: OK.

Gary Phillips: … hospital characteristics in the model because that's like – that was kind of a prioritizations. We were just going to use basic level characteristics.

And I think the other reason that, as far as generalizability, you know, if you start putting some hospital characteristics, I'm not saying like unique to New York and maybe the model couldn't be used out here in Ohio that or something like that. You know, there is a possibility that you put variables in the models that aren't generalizable to other parts of the country.

Male: The only thing I wanted to ask you about just our desire was there's – this is the first sepsis specific risk model and there – our intention was to create a big – a clean, a model based on justification characteristics and demonstrate that that model is responsive to interventions. And that was really the intent of building this risk model.

Melissa Marinelarena: Hi, and this is Melissa from NQF. I just want to provide a clarification with you at that risk adjustment. We do ask for the patient that is based on patient factors. If you were interested in the facility characteristics, sometimes they could be stratified by patient, you know, by hospital size or usually we'll see it by teaching, non-teaching hospital, by bed size, but the risk model should include patient factors.

Christy Skipper: Are there any other questions on the risk model itself?

Pranavi Sreeramoju: Is there published paper highlighting these data? I mean, there are a lot of data on the measure documents that have been shared with us, their peer-reviewed articles showing these data.

Foster Gesten: We're working on it.

Pranavi Sreeramoju: OK, OK. Fair enough. They take a long time.

Gary Phillips: Are there other questions that were asked of me before I was able to get on, enter into the building …

Foster Gesten: I don't think so, Gary. There may be others as we go through but I don't believe there were any. Thank you.

Gary Phillips: OK.

Christy Skipper: All right. So it sounds like we've covered some of your questions about the evidence for the measure and related to validity. So are there any questions or comments for the developer related to reliability? And you can see on your screen in front of you the measure worksheet which sort of – which summarizes the information the developers submitted.

OK, hearing no questions at this time regarding reliability, we can move on to the next criterion, feasibility. All right.

Melissa Marinelarena: Oh, actually, Christy, before we will move on, if we could just go to the validity section. And Foster if you and your team could just provide some information on how you did your data element validity testing, because I have some questions on that. We provided the percent agreement but I noticed that there was inter-rater reliability testing as well.

So, I just wanted – my understanding was that the inter-rater reliability testing was for the medical instructors, the auditors, prior to them actually doing the data element validity testing from the hospital charts. Is that correct?

Foster Gesten: Let me – (Kathy), hopefully you have an open line, maybe you can talk a little bit about what we did relative to looking at the accuracy and validity of the reporting from the hospitals. Because remember that the data that we have is

self-reported, abstracted by the hospital, submitted to us. And the open question for us and for use of the measure is how reliable, you know, how reliably can these specific, you know, variables be reported and abstracted to us particularly given that, you know, hospitals understood that this is in the context of, you know, public reporting initiative.

(Kathy), do you have an open line?

(Kathy): I think I do. Can you hear me?

Foster Gesten: Yes.

(Kathy): Yes. OK, great. I think, first of all, just kind of avoid some of the confusion that I just see based on the comments back. I think that we went to a much more granular level on the data that we submitted here when we looked back at some of the examples that were sent to us. What was noted is that the majority of folks who submitted their measures when they went to this level of the data they simply reported their kappa scores as being within acceptable ranges within the literature.

So, we went sort of a level below that and gave you a chart that gave the validity scores, the percent agreement as well as the kappa scores as opposed to that simple one that other folks gave that simply said that our raw data elements met the acceptable standards in the literature. So, I sort of give you that broad overview.

But that – but sort of the bigger picture of understanding the data was what Foster was speaking about, which is we could simply have accepted all of these data from the hospitals and simply accepted that as a fact and went forward with building our model. But we were very cautious. And I wanted to make sure that the data elements that we were putting into the model itself were in fact data that we could validate.

So, we went the next level and we audited a certain percentage of all of the medical records that were submitted to us, or all of the data lines that were submitted to us. Requested the medical records and actually validated each

data element that was submitted to determine that we could in fact find the answers submitted by the hospital in the medical record.

And for that, we have, in fact, reached the acceptable standards for ensuring that the data was accurate enough to then go forward and build the model for which you have all of your sensitivity scores in your larger tables that give you some indication of how the model performed given its small sample and then its application to the larger population.

So I'll stop at this point if there are more questions or more details that we can provide on that?

Pranavi Sreeramoju: No, I just have a comment. And this is Pranavi. You know, I envied New York state sepsis initiative, it was down at the state level. I think the data collection process was very – set up very well. And there are – the results there for everyone to see.

You know, we can't ignore the organizational characteristics or even the impact of having a state-based cohort like this. Because peer comparisons and peer interventions and local state-level interventions, these certainly have an impact in addition to the individual patient level improvement.

So, I mean, to extrapolate this to other states and other hospitals, not at a state level but at a hospital level or even at an individual level, I think we lose some of the organizational attributes and the state-level improvement initiative, the state-level attributes.

So I just hope that as these data are going to be used, I mean, they're suspected to be used to move some for pay-for-reporting towards a pay-for-performance. I hope there will be an intermediate phase there. It will be tested in other states and other groups of hospitals.

And by the time, hopefully, we'll have data from New York in terms of how the model performed in safety net hospitals versus academic hospitals versus V.A. hospitals, et cetera. So, that will be very helpful. And that's just a comment, it's not an ask or anything.

Christy Skipper: OK. Thank you. So we have a few more minutes to finish discussing mortality before we move on to the sepsis bundle. Are there any questions or comments regarding the feasibility or use and use or additional comments regarding the usability and use of this measure?

Pranavi Sreeramoju: I think the only thing with feasibility is, again, if hospitals were to set up a data collection process that are on these variables it would have to be manual. I don't see any of these – I mean, I don't see these variables being used – being collected electronically through EMR.

Some of them most but some are not and they do require extensive validation. So but that – that the – I mean I don't see that this is more laborious or less feasible compared to other measures that we collect data on, but it will impose original data collection bottom on hospitals.

Jamie Roney: This is Jamie Roney and just a comment on your comment. We're already having the mainly obstruct, those of our core measure data. Anyway, we found it very difficult to capture fluid resuscitation the EHR. So, I don't know that it maybe anymore laborious than those of us that are diving into the clinical record already looking at sepsis so closely.

Pranavi Sreeramoju: OK.

Laura Evans: This is Laura. Can I ask a quick point of clarification? So for talking just about the model itself, the risk adjustment model being used as a quality measure. They were talking about collective data elements that are included in the model, not the remainder of the New York State database that each side had to submit, is that correct?

Foster Gesten: You are asking – you ask …

Laura Evans: I guess yes. So – I mean we're talking about for this quality measure. It's the model – it's the variable than the model not …

Foster Gesten: Correct.

Laura Evans: … not the bundle adherence and other elements that we submit …

Foster Gesten: Correct.

Laura Evans: The New York State hospitals submit.

Male: Correct.

Male: Correct.

Laura Evans: OK.

Foster Gesten: I mean, as you know, New York hospitals are collecting other information responses to their rest initiative as well a, you know, responses to SEP-1 for CMS. But for purpose – for purposes of trying to do exactly what, you know, what Pranavi was talking about which is the use of that across state wide and across institutions and so on. It's really – it's a variable specifically that are in the model that will need to be collected. And they maybe a subset of variables that are collected for other purposes but they may not, you know, some of them maybe unique.

Laura Evans: And I would just add a comment to that – that I think that does improve the feasibility and usability of this because it is in general our variables that are less complicated to collect and some this fluid resuscitation variables that we have to collect for other purpose.

Foster Gesten: Yes.

Laura Evans: But there – most of the variables in the model a little bit more straight forward.

Foster Gesten: Yes. Thanks for making that comment. I think, you know, we were very concern about feasibility not only of the model but also there entire initiative. So we – you know, there are trade-offs intentions that we have between the potential list of clinical variables that we thought might be interesting in a model. And the absolute, you know, minimum that we though we needed.

And, you know, I think it was – I think it ended up in reasonable compromises of these two. But, you know, certainly when we look at the literature or look at, you know, Apache scores or other kind of risk scores, there are a lot more

clinical variable that might have been interesting to collect, might potentially contribute to the model. We didn't know if the time we're collecting, you know, what kind of performance we would get. We're ultimately – we're pretty happy with – where it ended up. But we had feasibility as an important characteristic of the data that would be in a model as we – as we were thinking about this.

Laura Evans: Thank. That's helpful.

Christy Skipper: Great. Are there any other questions or comments for Foster and his team?

Pranavi Sreeramoju: So, I have a question. So mortality, right? So this is in-hospital mortality, correct?

Male: Yes. That's correct.

Pranavi Sreeramoju: OK. And then …

Male: Yes.

Pranavi Sreeramoju: So how did New York approach coding standardization because sepsis is based on coding, right? And clinical documentation vary. So coding actually depends on clinical documentation. So how – how did New York State standardize coding process?

I mean I'm here to check also but I have – I'm really looking at this as a future type of performance measure and how the hospitals that I'm intimately involved with will perform and the greater good, how will hospitals handle it. So I'm really thing thinking at it – thinking about it from that perspective. So, do you have any …

(Crosstalk)

Foster Gesten: Two things, one I would say that – I'll answer it in two different ways and the full answer to your question is probably take more time than we've been – we have right now. But, one thing I would say is that while we have concerns despite what I'm going to tell you that what we did – how we address this, the fact that the risk adjustment, you know, is in part any answer to trying to, you

know, trying to address the issue of what happens if one hospital includes if you are the people who are the less severe spectrum of severe sepsis than another one.

And so, you know, the "severity variables" as well as comorbidity and other demographics is they're impart to mitigate or to try to address some of that regardless of what we – what are instructions were related to finding cases, right?

But in terms of – to more directly answer your question into – to try to give you a distinct answer, we told the hospitals for this initiative, which included the data in which we built the model to use the current international definitions to use both clinical and administrative data to find patients to screen them, to ensure that they met the clinical definition of severe sepsis and septic shock at the time. And we encourage them to, you know, to do these using concurrent or retrospective databases as well. We did not, for example, say only give us the patients who are coded, you know, X using ICD-9 or ICD-10 codes.

And in our audit process we also look to ensure that the patients that were in the data set, part of the audit was to ensure that in fact they met that – they met those definitions and we also did some cross-validation with the clinical data with the administrative data that does have those codes to see which ones were included and which ones are not included. So again, really interesting question. Some of this crosstalk between codes and clinical data is the subject of one of the papers that were put in together.

Pranavi Sreeramoju: OK, thank you. That's helpful.

Melissa Marinelarena: OK, thank you very much. One last time, is there anybody else from the committee who has a question or comment?

Christy Skipper: OK, all right. Thank you. So everyone, I just want to note that we will have a period toward the end of the call where members of the public can ask questions. For now, we are going to move on to measure 0500, Severe Sepsis and Septic Shock, Management Bundle. This is a measure by the Henry Ford

Hospital and I want to announce that we do have one recusal from this measure, Laura Evans, and she will not be discussing this measure.

So, I would like to turn it over to the developers, Manny and (Sean), if you could please just give a brief introduction to your measure.

Emanuel Rivers: This is Emanuel Rivers. Can you hear me OK?

Christy Skipper: Yes.

Emanuel Rivers: Good afternoon everyone and thank for joining this call. This is the 10th anniversary after the first submission to NQF, measure 500 that represents the third submission this time around. And the concept of this sepsis measure began about 20 years ago after observing the sepsis mortality approach in 50 percent at Henry Ford Hospital.

As a result, a quality initiative began in 1997 that gave way for the concept of early intervention in sepsis have interest and highlight in the challenge of treating sepsis the same mortality of 46.5 percent was validated last year by an expert panel in JAMA. Clearly, sepsis continues to be a condition where the gap in quality of care is significant.

The following are some note where the statistics from the Center for Medicare and Medicaid Services. There are about a million cases of sepsis met at hospital annually and 250,000 are septic shock cases, so about one quarter. Sepsis is the most expensive reason for hospitalization and it contributes up to 5 percent for the total U.S. admission cost.

And as you know, the mortality varies anywhere from 28 percent to 50 percent and it account for more deaths in prostate cancer, breast cancer, as well AIDS combine. And it's the fifth call the total hospital days and it's the largest increase in emergency department visits from the year 2006 to 2011. And so in essence, it kills approximately 258,000 patients a year.

As the committee knows, the concept of goal-directed therapy was a combination of mini studies that created components of the composite 0500 measure. One must remember that at that time in 2001 there were no

standards for early identification of care for sepsis, so protocol-based care was not the standard of care. This change after the adoption of the early goal-directed therapy treatment principles by the Surviving Sepsis Campaign in 2004 and after a decade, the Surviving Sepsis Campaign publications in multiple protocol-based care trials have shown that this mortality benefit been reproducible.

Now, three recent trials must discussion about the elements of goal-directed therapy. These protocol-based studies reveal about the core treatment elements of early detection, risk stratification, (flu) or challenge, cultures, antibiotics, repeat perfusion are simply implemented and the mortality is less than 20 percent. So it's important to note that these trials did not question protocol-based care that is the foundation of NQF 0500, but only the component related to the fusion exam.

We made specific modification to the measure to include the research in the last NQF submission, so hopefully this is no longer an issue. The adoption of the measure by CMS in 2015 in hospital data collection has provided previous significant insight to sepsis care. The sobering news is that the mortality at septic shock is still 38 percent to 42 percent, in severe sepsis 28 percent to 32 percent, which is twice as high of the mortality reported in the recent process trial.

Needless to say, we still have a lot of work to do to improve sepsis care in NQF 500 as lead in the charge to improve sepsis care nationally. So since the measure was implemented, the good news is then the first three quarters of data collection. SEP-1 show the mortality benefit of about 8 percent absolute and the relative reduction in mortality of 25 percent when implemented and completed.

We've observe this mortality reduction with compliance rates of only about 50 percent to 60 percent and it is note worthy that 100 percent of hospitals successfully reported the measure in the hospital performance has been increasing each quarter.

In response to thoughtful clinicians' stakeholder feedback regarding this specification manual, we've also made multiple changes to minimize clinician documentation that decrease hospital extraction burdens. We are very impressed with the early result of the measure on improving sepsis care and believe that NQF 500 will continue to be a vital component of the national sepsis quality improvement efforts.

So on the behalf of the measure steward, Sean Townsend, Henry Ford Hospital and I, I would like to thank NQF and CMS for the past support of this measure. After 20 years we've still – we've made significant strides in improving sepsis outcomes and I dedicate it to continue in our efforts to save the lives of sepsis patients. Thank you.

Christy Skipper: Thank you Manny for that great overview of the measure and the update to the measure that you're going to be – that committee is going to be reviewing in the couple of weeks. We'd like to start up with questions or comments from the committee.

Amesh Adalja: Hi, this is Amesh Adalja at University of Pittsburgh. I just have a question. I think that, you know, Dr. Rivers' trial that really was a breakthrough trial that change the way sepsis has viewed. I just – will this been some recent trials ProMISe and ProCESS for example where they haven't shown that the bundle could be completely – I haven't been able to completely replicate those findings and I wonder if Dr. Rivers or somebody else may discuss, you know, how we could endorse this measure with ProCESS and ProMISe sort of hanging over our head where they weren't able to show that kind of benefit.

Sean Townsend: This is Sean Townsend. I'd be happy to start there and ask Dr. Rivers to chime in if necessary. I appreciate the question and I understand where you're coming from. Dr. Rivers alluded to the ProCESS, ARISE and ProMISe trials in the beginning of his comments. It should be understood that SEP-1 was fully revised, that is 0500, was fully revised in committee at NQF to accommodate the process trials findings.

Dr. (Yili), the lead author in the trial became involved in the discussions and we removed the elements as a requirement to check CVP and ScvO2, which

were proxies for early goal-directed therapy. Instead, we place in the option for a physical exam, but none of the other pieces are actually not part in the sepsis 0500 metric or actually not different from what happen and process it.

It's really very important, and I'd like to remind the committee that what should be understood about process, ProMISe on arise, is that at the condition enrollment, 100 percent of those patients got lactate check, blood culture before antibiotics, broad spectrum antibiotics. And on average, three liters of fluid or rather 30 ml per kilogram of fluid prior to enrollment in that trial.

So, to say that they're not protocol-based or they don't include the element in 0500 is incorrect. Most of those things are actually in there. The other pieces that are in 0500 include the repeat assessment for volume status in the repeat lactate, which are consistent with the new Surviving Sepsis Campaign guidelines.

Emanuel Rivers: And I will refer to you an article we published comparing the three trials in addition to the methodology issues, because those trials were not blinded, which the early goal-directed therapy was blinded to the ICU clinicians, which is very significant because over time data patient of principles in terms of sepsis management, remember, there are 10 years between the goal-directed publication when those trials were done and a lot of the care overlapped into the control group.

So, we published this in Intensive Care Medicine about a year ago and most recent study we published in critical care that's might also alludes to the methodology issues between the original trial and history trials.

Christy Skipper: Are there any other questions or comments for the developers of measure 0500?

Esther Babady: Yes. Hi, this is Esther Babady. I'm from Sloan Kettering Cancer Center. I think one of the more question I had was with the recent guideline that was just published. Are you going back and sort of looking at some of the definition that are included in this measure, you know, sepsis versus severe sepsis and so on.

Sean Townsend: Hi, this is Sean Townsend again. Thank you for your question. I appreciate the guidelines. I'm listed as one of the authors on the guidelines, so I'm very familiar with our use of the new terms that were proposed in the Sepsis-3 definitions two Januarys ago at the Society of Critical Care Medicine conference.

SEP-1 or otherwise known as 0500 will continue to use screening criteria based on source and the terms of your sepsis. This is because all the literature reviewed essentially in the new guidelines published in '16, even though we adopted some of the nomenclature of Sepsis-3 are all based upon that very same screening technique, so SIRS criteria, suspicion of infection and an organ dysfunction were criteria for enrollment and all of these trials ARISE, ProMISe and ProCESS, for example. Even though guidelines addressed those using the vernacular of Sepsis-3 impart, the methods in the trials themselves referred to the definitions that preceded them.

If you look at the Sepsis-3 definitions papers very carefully, the authors actually don't suggest that they are ready for prime time use. In fact, they suggest that the methods need to be validated empirically before they're adopted for public. Some of that validation has began, it's not complete and there's been questions raised in particular (Inaudible) and (chest) last year as to whether or not early detection is compromised using the new definitions.

So taking the authors at their word and around the concerns that early detection maybe missed, a decision was made to keep the old definitions as the screening criteria in SEP-1.

Emanuel Rivers: And if I may add just a comment is when you look at the definition of septic shock by those new definitions, they include a lactate greater than two with vasopressin therapy as defining criteria for septic shock but 30 percent of patients actually with septic shock that are vasopressin dependent won't generate lactate.

And so, that flaw in that, you know, and that definition obviously needs to be readjusted and that also has significant coding per se when it comes to defining sepsis via ICD-9 code – ICD-10 coding.

(Off-Mic)

Christy Skipper: OK. Any other questions or comments from the committee?

Emily Aaronson: Sorry, this is Emily here. Just to clarify, any questions in general or are we going to be going through the validity and reliability piece in a moment?

Christy Skipper: We can move right on to reliability and validity if there are questions regarding that. So, did you have a question about reliability or validity?

Emily Aaronson: Yes. And I'm actually also in the cohort. Somebody mentioned earlier that I'm not able to view the survey that we filled out, but I know that we submitted those and some of those questions were included in there. I can look through my notes to pull them up, but not sure if you have them, the results from the survey there.

Christy Skipper: Yes. We do have those. We do not have those pull out at this time.

Melissa Marinelarena: When did you submit your survey, Emily?

Emily Aaronson: That was yesterday.

Melissa Marinelarena: Yes. Was it yesterday after, like, noon?

Emily Aaronson: Yes. It was before the 6:00 p.m.

Melissa Marinelarena: Oh, OK. Then I was not able to – I didn't send those to the developer, so if you can just ask your questions to them.

Emily Aaronson: Sure. I can let someone else take the lead. I just pull out my notes.

Melissa Marinelarena: Sure.

Michael Lane: Sure. I can ask the questions. I'm Michael Lane from WashU. I recognized that I had Ed Septimus is on the line too. I have a question for the developers – excuse me, antibiotic choices detailed in the measure. I know – I thank developers for being responsive and recognized the need for antimicrobial

stewardship and at times limiting the use of broad spectrum antibiotics when a known causative organism is there.

What measures or processes do you have in place to reassess the appropriateness of the antibiotics that are used or included as a – being in compliance with the bundle? For example, there are some antibiotics that are no longer on market, like ticarcillin-clavulanate, or those where resistance pattern suggest they might not be optimal choices, including some of the quinolones.

Sean Townsend: This is Sean Townsend, again. Appreciate the question. We have tried to make the antibiotic selection options broad, very broad. And so – and we have to reform the measure set a couple of times to make sure we were responsive to our colleagues in the infectious disease community and some of their concerns.

And so in that regard, for example, a couple of things that were done were to permit if we knew that particular organism was causing the sepsis syndrome in this particular case and that sensitivity – provider susceptibilities are known that agent that the clinician rights is an acceptable agent and that exception was approved. Likewise in the case of C Diff. we – if that was a causative organism, we said susceptibility testing wouldn't be required and therapy will just oral vanc or I.V. metronidazole would be appropriate. So we try to be responsive to these concerns as they come in and try to evaluate each one at its own terms.

In terms of the question of updating the antibiotic tables to reflect things like medications are discontinued and so forth, that's absolutely goal of ours. In fact, the original antibiotic tables included a PO medication. Augmentin was listed as something that could be use even though the specification is clearly stated that I.V. antibiotics were only allowed. That was fixed in one of our revisions.

We've had a total of three specification revisions and we tried to remove anything that's off market, but we can be – there's sufficient alacrity in the

revision process that if we are made aware of these things, we are happy to remove them if necessary or to add things that ought to be added.

You asked the question about resistance patterns, the tables are sufficiently broad to allow facilities who know their own resistance patterns to select from agents that would appropriate at those facilities. Nothing should be not in their options.

Michael Lane: Excuse me. If there are things that we know and nationally there's increasing resistance too, you know, including them at a acceptable option, you know, if we know that resistance raise, you know, nationally are relatively high, you know, as a timeline clinician it leads to the assumption that, you know, that is an acceptable therapy choice and, you know, that levofloxacin and moxifloxacin are probably some of those good examples.

Sean Townsend: Yes. Mike, I think you – I appreciate what you're saying. There's a question of whether, you know, by suggesting an antibiotic and antibiotic table people take license to say this is the appropriate antibiotic in this case. And we have to strike – we've tried to strike a balance in the measure between calling for broad spectrum drugs, but at the same time not removing clinician judgment about what the appropriate drug is.

As you can imagine what this measure, we're always in the cusp of too much restriction on clinicians versus not enough restriction clinicians. And that's a – that is certainly a balancing act.

We're open to ideas around this if there were statement we've received, for example, we've work at IDSA on some accommodations. If there were particular suggestions that we wanted to put into this, we, again, can do that in the future versions of the specifications.

Michael Lane: Right. Well, I defer at Dr. Septimus, who I know he was on the line as well and has done some work in this area.

Ed Septimus: Well, thank you. It's a great discussion. I want to say that it's been – Sean has I think framed things quite well. We've collaborated with the IDSA. CMS has been generally responsive to some of the changes which had been

discussed. The only other change that we've suggested that I don't believe CMS is quite ready to act on, which actually the Surviving Sepsis Campaign supports and that's the concept of reassessment and de-escalation and how that would be measured. But that was the only other thing that I think Infectious Society had concerns about. But CMS has been responsive to cleaning up those tables.

Sean Townsend: And as regard – so this is Sean, again. As regard that Ed, the question of antibiotic stewardship, you know, this is – the question is, could we put it into the measure somehow? And one of the – as you guys all are aware in this panel, one of the key requirements, of course, to get approve that a body like NQF or adopted is that we have data around particular elements that we proposed and then we can demonstrate certain facets of the measure that it's reliable or valid or feasible or usable.

And it's almost to catch 22 of – it's never been in the sepsis measure before they have antibiotic stewardship as an element. So there is, therefore, no data upon which to advance the element. It would be probably helpful, especially – as long as we're – before the infectious disease committee if, you know, from the public we got an antibiotic stewardship measure, which exceeded sepsis and look at many conditions. And then, you know, you could have a standalone side SEP-1, but difficult to incorporate it when the evolving body of data has never had to begin with.

Not – now, that's not to say that we don't support antibiotic stewardship. We agree with the guidelines. It is well specified in the 2016 Sepsis Campaign Guidelines that there should be reassessment after an appropriate period and tapering down of drugs to – those are in appropriate therapeutic class, et cetera, for the suspected organism. And that's all support – I don't think you'd find it any of the stewards here in disagreement with that. It's just that it's been treated separate from this measure.

Ed Septimus: I think that's a fair appraisal.

Sean Townsend: Thank you.

Christy Skipper: Were there any other questions, comments on the reliability or validity of this measure?

Emily Aaronson: This is Emily here. I guess a couple of questions just in looking through the validity piece. Obviously, as we can see a high number of the items, including some of the more foundational items that the rest of the measure has predicated on, like the severe sepsis presentation time had fairly low validity scores. The time zero had a validity score, percent agreement of 36.8 percent.

And I know there's a hypothesis put forth that this maybe higher now that's some education has happened. I wasn't sure if there was any data to support that hypothesis or any other thoughts around the concern that that was only 36 percent agreement for something that really sort of sets the clock for the subsequent compliance, sort of how that can be justified.

Sean Townsend: Sure. Emily, thanks. This is Sean Townsend, again. Your question is (present) and I want to answer in two parts. The first one is to state that as I told you, we update this measure frequently. And so since it's been deployed in October of 2015, three different updates have gone into effect that – or more in fact that alter the specifications. So we're now at version 5.2A.

And this is a function of, you know, what happen when we left committee at NQF previously in the measure was approved after we were compromising with the process trial authors is that we went to CMS and wrote the large specifications. And, you know, it's a first time product when it first gets done and it required a lot of revisions. And so some of the things, for example, around antibiotics you're describing had to be done. And just on the criterion you're describing on severe sepsis present the number of pieces were revised over those quarters because we found the way that it was being applied, it wasn't appropriate. This leads to a problem when in comes to assessing validity in the field.

The abstractors at hospitals have become used to one version, but then at update occur and that update has to be accommodated and understood. And because it's a moving process, the consequence I think is that we're seeing some of those lower validity scores around some of those data elements. That

was included in the submission because we wanted to be upfront with the committee that we are paying close attention to this validity scores and those with lower particular agreement.

I would mention, though, as a second part to my answer that it turns out that data element level validity for composite measure is not the measure standard of choice. We should be looking at the performance measure score for validity assessment for composite. And we have that type of information at the level of correlation of process compliance with mortality. And so that's helpful, but it doesn't eliminate the need to answer for the data element discrepancies.

What we've done and ensured to protect hospitals from dangers of that validity where there are maybe discrepancies is we have suppressed any kind of public reporting or benchmarking of this information until such time the measure settles. We believe now with version 5.2A becoming effective in January that the big revisions that needed to be accomplished have been done and our efforts will be to allow that to settle and people to become use to the algorithms and then reassess validity as we go along. So I do hope and expect to see improvement and it's our commitment that hospitals wouldn't be penalized in any way or even have their data revealed in the interim.

Emily Aaronson: Great. That's very helpful.

Christy Skipper: Any other questions or comments on reliability or validity?

Sean Townsend: I would – if I could just comment. It's Sean Townsend, again. And you need to stop me if I take up all the oxygen in the room, please. But, the – one thing – so you might have been say base on my comments on what I did – just made about suppressing the public reporting of the measure would value it to the hospitals.

Hospitals themselves are able to see their own performance measure score on the quality net that CMS sponsors that only the hospital considers their own score. So at the present moment, the value to hospital is that they're able to measure themselves, against themselves overtime with the consistent

measurement tool measured by their own abstractors in that way, so they can judge their own progress as time goes by.

Woody Eisenberg: So, Sean, this is Woody. So there's no benchmark that's visible to this hospitals?

Sean Townsend: That's right. So they don't – well, I suppose now that you've seen the submission to the extent that it's public some of the deciles with the overall metric are now in the submission and it has the first time they've been semi-released to the public. But, nobody, you know, whenever they get their own score, they're not told what decile they're in versus the rest of the country.

Emily Aaronson: And I guess this is a question that may, you know, that scopes beyond just at sepsis more towards measurement in general. But, you know, looking at those median scores understanding that, you know, you have medians ranging from 31 to 41 percent, which seems, you know, in keeping with compliance on other studies that have look at these bundles.

You know, the question is if there's a precedent to have a measure, especially when we think about it beyond NQF, but certainly at the CMS level, which is recommending care that we have never really tried to apply to a 100 percent or we have never apply to 100 percent of patients. And so, raising this concern if there is confounding by indication or these unmeasured confounders that exist in this observational studies in which roughly 40 percent of patients are receiving these bundles and if the implication to – especially, I think about some of these smaller hospitals who, you know, you receive recommendation from CMS and you're going to do that every single time.

And if we know that that phase for the 30 percent or 40 percent of patients who have not received that amount of fluids or, you know, antibiotics and rest of it. Does that – is that sort of make sense? I guess, you know, we obviously are in absence of any randomized trials where are 100 percent of patients have received this even though CHF studies that were referenced in the new appendix, none of those sort of said, let's take all comers and give them this

fluid and see how they do. It was really – when the physicians chose to give them the fluid, they did OK.

Sean Townsend: I'll try to unpacked your question and tell me if I get it wrong or I can give you better detail.

So, I guess I want to start with what Dr. Rivers begin at the top where he told you about the ProCESS, ProMISe, and ARISE trials. And that in both the control group and in the intervention groups, mortality was roughly the same, so finding of about 18 percent more mortality. And this really remarkable actually that that was consistent across three – essentially three countries but there were others involved in three large randomized control trials.

But, what is consistent in all those trials is that the majority of the elements that are in SEP-1, those patients received a 100 percent of the time. So you're asking about, do we have evidence support happens when you start to approach 100 percent compliance? And these patients were randomized to the trial. They were randomly selected and it was turned blinded fashion.

They received lactate as initial check, blood cultures before as antibiotics, a broad spectrum antibiotic. And on average, 30 ml per kilogram of fluid before entry.

We know they were in a randomized control trial with study coordinator etc cetera. So, I think they got frequent reassessment which is part of SEP-1.

And, in terms of repeating lactate, I have to look at the supplemental appendixes to know how often that was done. But, at least to this first five elements through fluids, those were all done at a 100 percent of the time, that achieve some mortality rate in three trials across, you know, continents of 18 percent consistently.

Now, what we look at – we looked at – when we look at for SEP-1 data we have, we see that mortality and those who failed the bundles about 30.4 percent. And those who passed the bundle it's 20 some odd percent, 21 percent perhaps.

And so this 8.5 percent reduction in mortality between the two groups and that 20 percent gets pretty close to the ProCESS, ProMISe, and ARISE if you do all the elements of the bundle. So it's really persuasive evidence that higher is better.

To your – another part I get fascinate and I think your question was this concept that, you know, we've never seen that though – when we actually apply to the real world outside of a randomized control setting getting close to a 100 percent. And that's true. The observational trials have all shown bundle compliance in that range 30 to 40 percent in general.

So, it does identify that there's a gap. And it will be very valuable to us once the measure stabilized and we can say, these are the deciles of performance. You know what, the best anybody can do and it looks like this is the case right now in the country, about 70 percent compliance, 60 to 70.

And hopefully people will then be able to judge themselves in comparison to that and not be striving in cases where people – maybe you should be deviate from the bundle for a 100 percent.

Emanuel Rivers: And if I may add one of the frequent issues is fluid management strategies.

Emily Aaronson: Yes.

Emanuel Rivers: We look at all these trials the average amount of fluid that was given in the first six hours between was between five and six liters of fluid which is consistent across all the trials. And actually it was more fluid than our early (goal-directed) study especially in a control group.

So, one of the misnomers is that they got the fluids pre-randomization that it was given. And therefore, we see sort of a fluid conservative interpretation from these trials, but the actual fluid given to these patients were the same.

Emily Aaronson: Interesting.

Sean Townsend: It is. You know, and we struggled a lot. This is Sean again. We struggled a lot with the questions that we received from the field about what you want to

do for congestive heart failure? What do you want to do for renal failure patients?

Emily Aaronson: Yes.

Sean Townsend: Because when we look at the – when we look at, for example, in Dr. River's trial and he could speak to this more eloquently than I. The subset of patients that had congestive heart failure and who were treated with a sufficient amount of fluids did worse than those that did get appropriate amount of fluids.

And that we've seen another observational trials. We send new paper by (Lou) and …

Emily Aaronson: Yes.

Sean Townsend: … the American Journal of Respiratory and Critical Care Medicine, yes, it sounds like you've seen it. Looking at intermediate lactate values and their strongest signal where in the patients they've got those fluids. So …

(Crosstalk)

Emanuel Rivers: Even if with hear failure.

Sean Townsend: Yes, even if with hear failure.

Emily Aaronson: Right. Yes. Go ahead.

Sean Townsend: Oh, I just going to say I'm – I've struck, I've racked my brain consistently about – is there some kind of exemption I could give for some degree of heart failure?

Emily Aaronson: Right.

Sean Townsend: Is there some kind of exemption I could give for end stage renal disease? But we all, you know, it's really difficult when the – we don't have an evidence basis from which to pick those things and none of the evidence says they do worse, we just haven't seen that yet.

Emily Aaronson: Right. And I think that's a struggle, you know, somebody that's leading this sort of sepsis related effort at our institution. Our cardiologists are deeply, deeply concerned about the fluid piece of this measure. And it has been a real struggle.

And I think what they've, you know, try to really articulate is that, although we don't have an evidence based to clearly, I think, demonstrate what a reasonable exemption could be or who that population is. You know, we don't really have a strong enough evidence based. They're concern is that we don't have a strong enough evidence based to really say that it's safe in a 100 percent of those patients.

And so, you know, obviously, our concern is always that if we put our guidelines that people will follow them. And we want to be sure, of course, that, you know, if that's happening that's a safe choice.

And so, that in the group of heart failure patients that in these observational trials are not getting the fluids. Is there some reason that that choice is being made? And are we setting up frontline clinicians for some degree of failure if we recommending that they give this amount of fluids to all patients?

That's, you know, just some inside – sort of what the frontline discussions are at our institution around the fluid piece. And, you know, they took a pretty deep dive through the evidence which most of – which was cited as you said and the recent update that we got.

But again, looking through this, the heart failure specific studies really having gotten to that randomized phase yet, does really open observational. So I think that's a little bit of a struggle.

Sean Townsend: So likely, we're in – we worked with antibiotic – we worked with the (I.D.) community around adjusting antibiotics, you know, we are open to receive suggestions like, for example, I recently considered – well, what if we said – what if we use some classification of heart failure like New York Heart Association? Instead, class four, there's a lot of not to do this, and of course, I haven't had people ask me about class three and so on.

We did put an exemption for VAD patients and that should be obvious why. But, you know, we've all have the renal failure patient, the end stage disease patient come to the emergency department having missed dialysis for couple days and they're probably volume overloaded but they're still septic in front of us and hypotensive.

In my practice, I give them the fluid. It's an end of one in an anecdote. But I can tell you overall – I did this, I can't put the specific to heart failure. But what I can tell you is that in our dataset that we've accumulated over the first three quarters, mortality in people who failed the measure as I allude to you is 30.4 percent.

Mortality in the group that failed the fluid measure was 37.7 percent. So, you know, there's signal there that this remains important, how to accommodate these gray areas is – has for now SEP-1 has been silent on it.

Emanuel Rivers: And Manny River again. You go back and look at the three trials ProCESS, ARISE, and ProMISe, the intubation rates were on the order of 25 to 30 percent which were 1/2 that of the (goal-directed) study. So, these patients got 5 to 6 liters of fluids. They got intubated less.

So, if you look at that association, you could say it perhaps that it may not be as deleterious as we think. And we look at our sub-analysis and show that people – and hemodialysis actually got the similar amounts of fluid, but the intubation rates were actually lower than patients who got less fluids.

So a lot of its counterintuitive but like you say perhaps the data needs to be more, you know, specific in term – in regards to those patients, but the observations there that they know just as good as other patients.

Sean Townsend: The other piece that I'll add to this since I appeal this question from audiences often not even this about SEP-1 but about sepsis in general because the guidelines continued to call for this 30 mls per kilogram. You've got to think about what – well what is a harm? Isn't intubation a harm? Or is it something that is part of care? And that depends on the patient I supposed and very frail or elderly person may not tolerate that intubation.

But rarely, is that the problem? Typically, especially and that becomes the volume issues. If you intubate somebody over food administration, you can get that fluid back off again and get them off the ventilator. If it's just fluid, if it's just fluid, they got them intubated.

Christy Skipper: Are there any other comments or questions on the scientific acceptability of the measure? If not, we'll move on to feasibility and usability in use. OK. Are there any questions or comments about the feasibility of the measure? OK.

Emily Aaronson: So if no one else has any question.

Christy Skipper: Go ahead.

Emily Aaronson: Sure, sure. And thank you again, that explanation is incredibly helpful around the food piece. And I think the insight around the intubation has not, you know, not necessarily looking at that as a poor outcome is very interesting.

The piece around the exclusions sort of (drums) out to me that when we were looking at the total number of excluded cases, you know, the vast majority of those were that severe sepsis wasn't present. I think it was 70, you know, in the low to mid 70 percent of the exclusions representing a fairly reasonable number of cases.

Now, I know our abstractors here, the feedback that we've gotten from them is that it actually takes a lot of abstraction to get to the point where you can determine that it is not – that for sepsis isn't present. And so, I think just some other reflection and the explanation of that was little concerning which was sort of implying that it reduces abstractor burden being able to exclude all those cases but I think that, you know, our experience at least just that the abstractor burden of this measure overall as we all know is pretty significant. But the yield of reviewing so many cases that end up being excluded.

I guess the question is if there was any thought to a way to make that piece a little tighter, to not have that amount of burden for such a low yield.

Sean Thomson: It's Sean again. It's a great question. And I think we recognized that these metric is very complex for hospitals to abstract. I think, you know, one of the criteria that's not in the NQF criterion but is alluded to I guess in performance gap or opportunity is, it's the question of burden of disease, you know, for the country.

And so, all the questions are on usability and feasibility. I'd like to put them and read along the side column, how important is it for us to be taking care of this disease and focusing resources in the hospital on it. So the statistics that Manny cited at the beginning, you know, the million cases annually 250,000 cases that shocked cost to 5 percent all U.S. admissions, et cetera. I think clear – put that in the certain vein as we have this discussion of that said.

There's a double-edge sword in here and many of you I think on the panel will appreciate this. The question of people would talk – when they talk about sepsis, there's an expression of an expanding denominator because we're getting better at screening for patients and how that therefore can found metrics around the sepsis.

And there's this question of whether coding therefore because of this earlier and better detection really is a way that we should be comfortable looking at a sepsis metrics. So for example, you just considered the New York State metric, the basis for SEP-1 and 2 is also initially coded simple as patients that has either simple sepsis, severe sepsis or septic shock. ICD-10 will recognizes those classifications and so that they are coded into those groups.

Because we wanted to generate the valid measure, we didn't want to rely on coders not trained in clinical activity and, you know, that that set of patients within expanding denominator would define when it is SEP-1 exclusively. So there, you get to the root of your question which is the abstractors then have to do to work of taking that code instead of patients, and then making sure that they meet severe sepsis or septic shock criteria as clearly specified in the specifications of the measure.

So yes, we toss out a lot of cases because the coders said something. But yet we can't – with clinical criteria reach the same conclusion. And you can see the balance in there already.

This is fascinating question really. And so, you know, I think the third part to your question was, it there something we could do to reduce that? And I have – the reason thought that I have had and I'm going to just speak …

Female: And quick question, first.

Sean Townsend: Sure. I'm sorry, I thought I heard an interruption. But a recent thought I had and I'm just speaking extemporaneously now is what might be valuable and what might also assist with that data element validity, the question we had earlier, is that as we go forward, if I could say of the sample you have to collect, so let's say at a month you're doing 40 cases. If the abstractor was given a 10 percent discretion rate to say the criteria in conflict, something is – that charting is in conflict.

Emily Aaronson: Right.

Sean Townsend: The notes don't make sense or I can't quite figure out if this was the intention, to just toss that case. So I'm not going to include it in the sample, they've got that much discretion and that might be a valuable thing for the future, so that's a consideration that we could certainly make.

Emily Aaronson: Yes, that's a very interesting option because I know that our coders, you know, involves in our quality work on our abstractors for all of our different hospital measures and there is no question that this measure has cost the most strife amongst our abstractors, and just the amount of digging that they have to do and just not feeling confident that they have the right answer because these data elements are recorded so variably in so many different places in the chart by every different clinician in a different way.

And so, you know, we're certainly hearing a lot of concern that they're not capturing this appropriately. And so, I think that that's really interesting potential solution for that is that if they're facing a chart with particular ambiguity they can toss it.

Sean Townsend: Yes, I'm really – I'm partial to it. I'm getting, you know, I just came – I'd be frank with you, I just thought of that about 5:30 this morning when I got up, so something for us to work on.

Emily Aaronson: Sure.

Ed Septimus: Is that Pacific time, Sean?

Sean Townsend: Yes, it is. So you are all ...

(Off-Mic)

Sean Townsend: ... for a long time.

Emanuel Rivers: But also I wanted to add that some of this excluding criteria actually favors the mortality in reality because transfer from other hospitals can be very significant in our institution up to 15 percent of septic shock comes also at the hospitals, these are exempt. But as we know effective resuscitation can be highly beneficial up to 12 hours after the onset at septic shock and these patients are actually excluded.

And you look at septic shock expired in less than six hours, up to 20 percent of patients would die within six hours of their septic shock episode, those patients are excluded. So you look at the true reality of life and say, well, these 30 percent, 40 percent mortality that we're seeing now is probably much higher than that because we're not including these patients.

Sean Townsend: Yes, and I'm not trying to be a little funny here when I say this, but we're trying to resemble the new large trials. I think you might – you might remember in process that about half of the patients were excluded before it was analyzed.

Emily Aaronson: Right.

Christy Skipper: OK. Moving on to usability and use the final measure evaluation criteria. Do the committee have any questions about the usability of the measure or comments?

All right, hearing none …

Sean Townsend: Can I actually – this is Sean. Can I make a comment?

Christy Skipper: Sure.

Sean Townsend: And may be still think of something about usability and use for the meantime. But back on feasibility for a second, we did include – so if you framed it the way I asked you to look at it again, you know, burden versus – data burden of collection versus disease burden in the country and the causes to us. One thing I can point to that I think is useful is that of the 3,000 plus acute care hospitals in the country subject to this measure. In every quarter since it went live, 99 percent of the hospitals have been able to report all the data elements. So it's – it is empirically, it's feasible, but it is also an amount of work to be done.

Emily Aaronson: Yes. And I would say the one caveat to that is when we look back at those percent agreements, I think, you know, we certainly submit every month, but every month I am at some points speaking with our abstractor about her sort of discomfort with that, so, you know, validity of the data that she's abstracting, so.

Sean Townsend: I run a – you know, my job and my day job is I'm an administrator at a hospital here in San Francisco and I've run the quality department. And so, I was called abstractors and I know them all personally. So I hear my share of stories.

Emily Aaronson: Yes, yes.

Christy Skipper: All right, if there were no questions or comments from the committee on usability or any other general comments about the measures, this measure 0500, we'll move to public and member commenting. Just want to give one final chance for any final questions before we move to commenting?

All right, thank you. Operator could you please open the line to hear if there are any member or public comments?

Operator: Thank you. To make a comment, please press star then the number one on your telephone keypad. We'll pause for just a moment.

And there are no public comments for this time.

Christy Skipper: All right, thank you. So that concludes our call for today. And I want to thank everyone for participating. We look forward to seeing you all at our in-person meeting in two weeks where we will discuss and vote on all the measures submitted to this project. So thank you all for your time and have a good afternoon.

Emily Aaronson: Thank you.

Male: Thank you.

(Crosstalk)

Male: Thank you.

Male: Thank you.

END