NATIONAL QUALITY FORUM

Memo

TO: Consensus Standards Approval Committee (CSAC)

FR: Melissa Mariñelarena, Senior Director; Christy Skipper, Project Manager; Mauricio Menendez, Project Analyst

RE: Infectious Disease, 2016-2017

DA: July 11, 2017

CSAC ACTION REQUIRED: The CSAC will review recommendations from the Infectious Disease project at its July 11-12, 2017 meeting and vote whether to uphold the recommendations from the Committee.

This memo includes a summary of the project, recommended measures, and themes identified from and responses to the public and member comments.

NQF Member voting on these recommended measures closed on June 28, 2017.

Accompanying this memo are the following documents:

1. **Infectious Disease Draft Report.** The draft report has been updated to reflect the changes made following Standing Committee discussion of public and member comments. The complete draft report and supplemental materials are available on the project page.
2. **Comment Table.** Staff has identified themes within the comments received. This table lists nine comments received during the post meeting comment period and the NQF/Standing Committee responses.

BACKGROUND

For this project, the 16-member Infectious Disease Standing Committee met during a one-day in-person meeting to evaluate nine measures against NQF's standard evaluation criteria. The Committee evaluated four newly-submitted measures and five measures undergoing maintenance review. All nine measures were recommended for endorsement.

DRAFT REPORT

The Infectious Disease Draft Report presents the results of the evaluation of nine measures considered under the Consensus Development Process (CDP). All nine measures are recommended for endorsement.

The measures were evaluated against the 2016 version of the measure evaluation criteria.

	Maintenance	New	Total
Measures under consideration	5	4	9
Measures recommended for endorsement	5	4	9
Measures withdrawn from consideration	7	0	7

CSAC ACTION REQUIRED

Pursuant to the CDP, the CSAC is asked to consider endorsement of nine candidate consensus measures.

Infectious Disease Measures Recommended for Endorsement:

- 2082 HIV Viral Load Suppression

Overall Suitability for Endorsement: Y-16; N-0

- 3210 HIV Viral Load Suppression (Legacy eMeasure)

Overall Suitability for Endorsement: Y-16; N-0

- 2079 HIV Medical Visit Frequency

Overall Suitability for Endorsement: Y-15; N-0

- 3209 HIV Medical Visit Frequency (Legacy eMeasure)

Overall Suitability for Endorsement: Y-16; N-0

- 2080 Gap in HIV Medical Visit

Overall Suitability for Endorsement: Y-13; N-0

- 2083 Prescription of HIV Antiretroviral Therapy

Overall Suitability for Endorsement: Y-12; N-0

- 3211 Prescription of HIV Antiretroviral Therapy (Legacy eMeasure)

Overall Suitability for Endorsement: Y-12; N-0

- 0500 Severe Sepsis and Septic Shock: Management Bundle

Overall Suitability for Endorsement: Y-10; N-4

- 3215 Adult Inpatient Risk Adjusted Sepsis Mortality

Overall Suitability for Endorsement: Y-11; N-5

COMMENTS AND THEIR DISPOSITION

NQF received nine comments from eight organizations (including five member organizations) and individuals pertaining to the general draft report and to the measures under consideration.

A table of comments submitted during the comment period, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the Infectious Disease project page under the Public and Member Comment section.

Comment Themes and Committee Responses

Comments about specific measure specifications and rationale were forwarded to the developers, who were

invited to respond.

The Standing Committee reviewed all of the submitted comments (general and measure specific) and developer responses. Committee members focused their discussion on measures or topic areas with the most significant and recurring issues.

Theme 1 – Antibiotic Administration

Measure 0500: Severe Sepsis and Septic Shock: Management Bundle received four comments concerning the administration of antibiotics in patients suspected to have sepsis or septic shock. One commenter noted that the 3-hour time window for antibiotic administration might lead to unintended consequences like antibiotic overuse. Two commenters questioned whether antibiotic administration rather than early goal directed therapy (EGDT) had a larger effect on patient survival rates as suggested by Kalil et al. (2017). One commenter suggested reducing the antibiotic administration time from three hours to two.

Developer Response to Comment ID 6678: We understand that measures can have unintended consequences for patients. We remain convinced that in severe sepsis (please refer to clinical criteria used for the measure) and septic shock the risk of mortality is so high that is critical to provide early and broad antibiotic therapy. The failure to provide a proper antibiotic for a patient with severe sepsis and septic shock carries a much higher risk than the potential harm of providing a single dose of a broad spectrum antibiotic to a patient who turns out not to have severe sepsis or septic shock. This in by no means precludes the clinician's responsibility to judiciously use IV antibiotics in a way which upholds the standards of antibiotic stewardship. The 2016 Surviving Sepsis Guidelines, endorsed by the Infectious Disease Society of America speak to this question: "The rapidity of [antimicrobial] administration is central to the beneficial effect of appropriate antimicrobials. In the presence of severe sepsis or septic shock, each hour delay in administration of appropriate antimicrobials is associated with a measurable increase in mortality." When mortality already approaches 18-40% in shock states, it is unacceptable to suspend antibiotic administration pending further studies. However, in the event an antibiotic is given inappropriately in non-sepsis states, the guidelines also recommend, "Given the potential harm associated with unnecessarily prolonged antimicrobial therapy, daily assessment for de-escalation of antimicrobial therapy is recommended in patients with severe sepsis and septic shock."
Developer Response to Comment ID 6680: The PRISM investigators have reported results in the New England Journal of Medicine (NEJM) entirely consistent with the prior Process, Promise and Arise trials, also published in NEJM. Little information is imparted by PRISM that was not already known from these previous trials. In fact, PRISM derived all data from the prior trials. We emphasize that SEP-1 is consistent with the conclusions of these trials and does not require an invasive method of patient reassessment. In regards to Dr. Kalil's publication in Critical Care Medicine, "Early Goal-Directed Therapy for Sepsis: A Novel Solution for Discordant Survival Outcomes in Clinical Trials," the major conclusion was that:

> "[S]urvival discordance was not associated with differences in early goal-directed therapy bundle compliance or hemodynamic goal achievement. Our results suggest that it was associated with faster and more appropriate antibiotic co-intervention in the early goal-directed therapy arm compared with controls in the observational studies but not in the randomized trials."

While we may dispute the methods and analysis used to reach this conclusion, we again underscore that SEP-1 does not mandate an invasive reassessment but does require early IV antibiotic administration. Thus SEP-1 is consistent with the Kalil publication in its approach.

Developer Response to Comment ID 6806: The developers will take the Armstrong Institute's helpful suggestion under advisement and model data to understand how a 2 hours standard would affect the performance characteristics of the measure. We agree that earlier administration of antibiotics is the preferred approach.

Committee Response: Thank you for your comment. The Committee agrees that monitoring for unintended consequences is an important part of measure development and implementation and quality improvement

programs. The Committee recommends the developers continue to modify the measure specifications (as needed) as the evidence in sepsis and septic shock continues to evolve.

Theme 2 – Level of Evidence

Measure 0500: Severe Sepsis and Septic Shock: Management Bundle received two comments noting the varying level of evidence for the different components in the measure composite. The comments suggested that the level of evidence supporting repeat lactate, fluid reassessment, and/or physical exam is not equivalent to antibiotic or IV fluid administration. The comments state that the components should not be weighted equally in the construct of the composite measure. One commenter also recommends a simplified 3-hour bundle without the repeat lactate and physical exam component.

Developer Response to Comment ID 6708: We will take under consideration the suggestion to weight elements in different ways than presently weighted. It is important to note however that as a matter of process for vetting measures, they must be advanced on the basis of accumulated data and evidence. SEP-1 continues to show a high association with reduction in mortality with the individual specified elements as documented in the submission which justified the weighting. As the submission shows, there is a large separation in mortality between those who comply with the elements in total and those who fail any one or more than one of the elements. To understand and model a proposal such as Dr. Doerfler's will require analysis of the measure as a component measure, which necessarily means analyzing the data in a fashion for which it was not designed. We will discuss with CMS Dr. Doerfler's hypothesis regarding preferential weighting of certain data elements.

Developer Response to Comment ID 6810: The developers appreciate the opportunity to respond to the High Value Healthcare Collaborative (HVHCC) which has advanced excellent work in quality improvement and care of patients with severe sepsis and septic shock. We note that the HVHCC has indicated that SEP-1 endorses an "all-or-none payment approach." SEP-1 performance does not impact payment to hospitals or providers. As regards concerns that SEP-1's composite construction does not differentiate between importance of antibiotic administration and a physical exam element (cited as skin color), the developers would like to point out that, in the current specification manual and in this NQF submission, SEP-1 does not require documentation of particular physical exam elements. A provider may now indicate simply (within the allotted timeframe) that they have "performed a physical exam" without regard to a means or method, physical exam or otherwise, and pass this data element. In this regard, the developers would suggest that regular reassessment of a patient with septic shock as regards to perfusion status is as important as antibiotic administration and supports the composite construction as advanced in this submission.

As regards the representation that "once a mature care model is in place, compliance with a 3-hr-bundle, had no impact on in-hospital, 30-day, 90-day or 1-year post discharge mortality between those receiving the full bundle vs not. This is consistent with the conclusions from the ProCESS, ARISE, and ProMISE trials," this claim is not an accurate representation of the cited trials. The 3 hour elements of care were required of every patient in the cited trials. All patients received the three hour elements (initial lactate, blood culture collection, and broad spectrum antibiotic administration) prior to randomization. Aside from this inaccuracy, it is unclear what the characteristics of a "mature care model" may be in the HVHCC's remarks, however generalizing that the 3000+ hospitals in the United States subject to SEP-1 have such a model is unsupported by any evidence to properly analyze that claim. Moreover, since HVHCC does endorse a "simplified 3-hour-bundle" it would seem to remain an important element of care. The developers do not believe that clinicians and hospitals may not define "innovative approaches to early sepsis detection" under SEP-1. One method to ascertain time zero that overrides all other methods is a provider's documentation of the time. In that regard, the HVHCC may use whatever method they prefer to set a time zero as long as their clinicians concur with the HVHCC's approach. The developers appreciate the HVHCC's suggestion to proceed with the 3 hour elements in SEP-1 and we assure them that these elements remain in this submission. As regards the representation that there is "no evidence" supporting repeat lactate assessment or a physical exam, the developers repeat that clinicians must only

document reassessment of perfusion or volume status by any means of their choosing. This practice, along with repeat lactate assessment do have a supporting evidence base in the 2016 Surviving Sepsis Campaign guidelines. Reassessment is a best practice statement under the GRADE evaluation framework and repeat lactate assessment has a low quality of evidence with a weak recommendation under the same criteria. We will take under advisement that these elements should be examined further in future iterations of the specifications and would welcome the opportunity to work with the HVHCC to model these approaches. We note that the measure does not apply to critical access facilities and is not active at this time in any pay for performance programs.

Committee Response: Thank you for your comment. The Committee discussed the varying level of evidence for the different components and agreed that overall, the updated evidence is consistent with the 2016 Guidelines for Management of Sepsis and Septic Shock. Additionally, the Committee recommended that the developer complete further testing on weighting the individual components.

Theme 3 – Scientific Acceptability

Measure 0500: Severe Sepsis and Septic Shock: Management Bundle received two comments questioning the reliability and validity testing results, specifically the patient level data element percentage agreement rates. The commenters also disagreed with the guidance given to the Standing Committee on evaluating composite measures. NQF criteria states that for composite measures, validity and reliability should be empirically demonstrated at the measure score level.

Developer Response to Comment ID 6803-6804: The Federation of American Hospitals has raised many questions that were previously discussed in detail in committee. We appreciate the opportunity to summarize these issues. While burden of data collection may be greater than for other measures in healthcare, this is more than counterbalanced by severe sepsis and septic shock's' burden on the healthcare system as the number one cause of inpatient deaths in the United States and highest cost condition for hospital admissions. Evidence that the SEP-1 measure drives quality improvement was provided at NQF. The initial three quarters of data analysis show that hospitals improved their performance from quarter to quarter. In addition, the analysis revealed a statistically significant finding that there is an approximately 8.5% associated reduction in mortality in those who comply with the measure versus those who fail the measure. As measure developers we have no data to comment on the quality of responses provided by Quality Net, but we will share the feedback with the Quality Net team. We cannot address the Federation's representations regarding the motives of other agencies to utilize the measure, but we have provided substantial data that SEP-1 meets the standards set out in NQF's measure evaluation framework. As regards validity, the measure met standards for assessing validity at the performance score level, which is the proper level of evaluation for a composite measure. Additionally, it is precisely for the Federations' argument of the limited sample size (303 cases) that the data element level validity cannot serve as a valid critique of SEP-1. In addition, the measure met all reliability criteria with statistically appropriate analysis using a signal to noise methodology. While the Federation states that the element level validity testing is more important than the performance measure score testing, under the NQF measure evaluation framework, that choice is an improper standard to evaluate a composite measure. We note the Federation misinterprets the Technical Expert Panel's remarks that "the individual components may not be sufficiently reliable independently, but could contribute to the reliability of the composite performance measure." First, this quotation refers to reliability whereas the Federation was addressing validity. Secondly, the principle that the individual elements could contribute to the overall validity at the level of the performance score is precisely the point of the Technical Expert Panel's comment. This rationale is why element validity is not the criterion for composite measures. Finally, the Federation has advanced no evidence to evaluate the claim that the measure does not meet the validity criteria set out by NQF at the performance measure score level.

The Federation has inadvertently misstated the evaluation criteria. Specifically, subcriterion 2d states that "[f]or composite performance measures, empirical analyses support the composite construction approach…" Nothing

in the Federation's remarks indicates that the composite construction approach is under question. As regards "missing data," the measure evaluation frame work considers this under subcriterion 2b7, which requires that the measure "analyses and identifies the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders and how the specified handling of missing data minimizes biases." This requirement is different than the scenario described by the Federation which equates the variation in element level validity testing with missing data. This equivalency is incorrect insofar as the analysis of the complete data shows that some degree of variation actually exists – if the data were "missing" a showing of variation could not be made. In this regard, the developers stand by their submission that all data is present and cannot be missing as part of the reporting requirements met by 99.9% of over 3200 acute care hospitals in three consecutive quarters of analyzed data.

To the extent that SEP-1 data elements may be contained in an electronic health record, the developers will take the Federation's excellent suggestion under advisement to consider, if feasible, the measure as an e-measure. The Federation implies that hospitals do not understand why they fail the measure given its composite nature. The developers believe, however, that by analyzing the point of failure in the measure framework, providers know exactly where weak spots are in their clinical care processes. For example, failure of the antibiotic element indicates a process in antibiotic administration that needs attention. In addition, the software provided by major vendors to hospitals to report the measure specifically categorizes the level of the fallout for facilities. The developers strenuously object to the characterization that the measure has limited value in improving patient care: in over 600,000 patients captured by the measure there has been an approximately 8.5% reduction in mortality in those compliant with the measure versus those who were not. For the population of submitted cases, this represents a potential lives saved calculation of over 7,500 patients in the first three quarters of data for the measure.

Developer Response to Comment ID 6811-6813: The developer's appreciate the opportunity to respond to the remarks of the American Medical Association (AMA) and clarify the operation of SEP-1. As regards to Dr. Pronovost's publication in the American Journal of Medical Quality regarding possible unintended consequences of quality measures, we note that Dr. Pronovost is the Director of the Armstrong Institute for Patient Safety and Quality which has recommended that SEP-1 be re-endorsed with the exception of tightening the antibiotic administration requirement from 3 hours to 2 hours. Please see the submitted comment of Dr. Matt Austin, PHD, on behalf of the Armstrong Institute for Patient Safety and Quality at Johns Hopkins University, which was also received during the post-evaluation comment period.

With respect to the specific concern that fluid administration as specified in SEP-1 may be harmful to patients with left ventricular systolic dysfunction, the AMA cites an opinion article authored by an emergency medicine resident. We note that this opinion is not representative of any clinical trial, observational or randomized, controlled or not. The opinion cites another article by Boyd 2011 which indicates that a positive fluid balance and elevated CVP are associated with increased mortality. The developers note that SEP-1 does not advocate for a "positive fluid balance" which refers to volume status over several days of care. SEP-1 is limited to initial resuscitation and the first 6 hours of care after presentation. In addition, SEP-1 does not advocate for CVP measurement, and certainly not an "elevated CVP." Another citation in the resident's article is Pudilo 2012 which reports frequency of myocardial dysfunction in severe sepsis and septic shock. The article by Pudilo actually points to a reason for proper volume resuscitation of patients: the well-known presence of severe sepsis induced myocardial dysfunction. The salient finding is not that the patients have intrinsic heart disease, but rather that sepsis has caused impaired myocardial function. In addition, Pudilo does not conclude that a 30 ml/kg initial fluid bolus as an initial resuscitation strategy in septic shock is detrimental to patients. This Pudilo reference does not support any of AMA's criticism of the SEP-1 measure.

On the broader concern about the potential risks of fluid resuscitation, we note that there is no published evidence from any randomized controlled trial which indicates that patients with severe sepsis and known congestive heart failure or renal failure who receive a fluid bolus for initial resuscitation do worse than other sepsis patients in terms of mortality. In fact, even the examination of the large trials on septic shock do not support this contention (EGDT 2001, Process 2014, Promise 2014, Arise 2015). In fact the only published evidence on the topic concludes that for patients with intermediate lactate vales of 2-4 mmol/L who receive the full fluid bolus of 30 ml/kg with congestive heart failure and renal failure have lower mortality than their counterparts without these co-morbidities. (See: Lui V et al. Fluid Volume, Lactate Values, and Mortality in Sepsis Patients with Intermediate Lactate Values. Ann Am Thorac Soc Vol 10, No 5, pp 466-473, Oct 2013).

The thrust of the AMA's comments on this topic regarding LVSD is that physician judgment should be preserved. The developers agree with the AMA that physician judgment is paramount and agree that providers should exercise their best judgment informed by the evidence when caring for sepsis patients. SEP-1 is not a prescriptive recipe for all patients with severe sepsis and septic shock, but rather a measurement strategy for processes of sepsis care. The developers fully expect that practitioners will do what is best in their understanding for each patient, which may result in deviation from SEP-1. Ultimately when sufficient data is amassed, best compliance, which takes into account those necessary deviations, will be known. In that regard, there is no expectation or goal of a 100% compliance with SEP-1. However, we emphasize that all current evidence suggests better mortality with higher compliance.

As regards to the concern that a precisely specified measure should account for how unplanned drug shortages could impact an individual hospital's performance and the concern that mortality varied with a shortage of nor-epinephrine, we note that SEP-does not require the use of any one particular vasopressor. We also note that SEP-1 is a process measure, not an outcome measure such as mortality. Finally, we note that it is likely that all hospitals would be affected by any shortage in a short period of time.

Turning to Dr. Kalil's publication in Critical Care Medicine, "Early Goal-Directed Therapy for Sepsis: A Novel Solution for Discordant Survival Outcomes in Clinical Trials," the major conclusion was that "[s]urvival discordance was not associated with differences in early goal-directed therapy bundle compliance or hemodynamic goal achievement. Our results suggest that it was associated with faster and more appropriate antibiotic co-intervention in the early goal-directed therapy arm compared with controls in the observational studies but not in the randomized trials." While we may dispute the methods and analysis used to reach this conclusion, we again underscore that SEP-1 does not mandate Early Goal Directed Therapy and SEP-1 does require early antibiotic administration. Thus SEP-1 is consistent with the Kalil publication in its approach. The PRISM investigators have reported results in the New England Journal of Medicine (NEJM) entirely consistent with the Process, Promise and Arise trials, also published in NEJM. Little information is imparted by PRISM that was not clearly know from the three other trials, and in fact PRISM derived all data from the prior trials. In any case, SEP-1 as specified is consistent with the conclusions of these trials.

AMA has stated concerns with SEP-1's measure performance characteristics. For its validity, the measure met standards for assessing validity at the performance score level, which is the proper level of evaluation for a composite measure. In addition, the measure met reliability criteria with statistically appropriate analysis using a signal to noise methodology. While AMA represents that element level validity testing is more important than the performance measure score testing, this is an improper standard to evaluate a composite measure under the NQF measure evaluation framework. Of substantial importance, we note that AMA misinterprets the Technical Expert Panel's remarks that "the individual components may not be sufficiently reliable independently, but could contribute to the reliability of the composite performance measure." First, the quoted principle that the individual elements could contribute to the overall validity at the level of the performance score is precisely the point of the Technical Expert Panel's recommendation not to focus on element level testing for a composite metric. Second, the AMA has advanced no evidence to evaluate the claim that the measure does not meet the validity criteria set out by NQF at the performance measure score level. In the evaluation of a metric, it would be improper to move the goal post mid-evaluation.

Regarding their other comments on validity, the AMA incorrectly cites subcriterion 2d. Subcriterion 2d states that "[f]or composite performance measures, empirical analyses support the composite construction approach…" Nothing in the AMA's remarks indicates that the composite construction approach is under question. In regard to "missing data," the measure evaluation frame work actually considers this under subcriterion 2b7, which requires that the measure "analyzes and identifies the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders and how the specified handling of missing data minimizes biases." This requirement is different than the scenario described by AMA which equates the variation in element level validity testing with missing data. This equivalency is incorrect insofar as the analysis of the complete data shows that some degree of variation actually exists – if the data were "missing" a showing of variation could not be made. In this regard, the developers stand by their submission that all data is present and cannot be missing as part of the reporting requirements met by 99.9% of 3225 acute care hospitals in three consecutive quarters of analyzed data.

In summary, the developers appreciate the AMA's comments and suggestions. We agree with the AMA that this disease which places an unacceptable burden in terms of deaths and expenditures on the healthcare system deserves a rigorous and robust measure. For this reason we are proud that SEP-1 (which has over 600,000

reported cases since inception) has an associated 8.5% reduction in mortality for those cases that were compliant with the measure versus those which were not.

We look forward to tracking the ongoing impact of the SEP-1 measure on sepsis quality improvement and will share our findings with all concerned stakeholders as they become available.

Committee Response: Thank you for your comment. The Committee encourages the developers and CMS to provide education and outreach, continued efforts to decrease abstraction burden and successive data element validation testing.

NQF Response: Thank you for your comment. NQF composite performance measure evaluation guidance (2013) states that validity testing is directed toward the inferences that can be made about accountable entities on the basis of their performance measure scores. For the purposes of endorsing composite performance measures, validity testing of the constructed composite performance measure score is more important than validity testing of the component measures. Even if the individual component measures are valid, the aggregation and weighting rules for constructing the composite could result in a score that is not a true reflection of quality (p. 12-13). Thus, the data element testing results can be considered, but the score-level results are of greater interest. Please note that the statement in NQF's composite report that states "the individual components may not be sufficiently reliable independently, but could contribute to the reliability of the composite performance measure" is actually referring to score-level reliability testing of the components, not data element-level testing of the components. In other words, it is possible that there is not sufficient signal for an individual component, but inclusion of that component will increase the reliability of the overall composite. The updated reliability and validity composite score level testing provided by the developer meet NQF criteria for composite measures as indicated during the in-person meeting and in the draft report.

NQF MEMBER VOTING RESULTS

ALL recommended measures were approved with 67 % approval or higher. Representatives of three member organizations voted; no votes were received from Health Plan, Health Professional, Public/Community Health Agency or Supplier/Industry Councils. Results for each measure are provided in Appendix A.

REMOVAL OF ENDORSEMENT

Seven measures previously endorsed by NQF have not been re-submitted for maintenance of endorsement.

Measure	Measure Description	Reason for Removal of Endorsement
0393 Hepatitis C: Confirmation of Hepatitis C Viremia	Percentage of patients aged 18 years and older who are hepatitis C antibody positive seen for an initial evaluation for whom hepatitis C virus (HCV) RNA testing was ordered or previously performed	In order to align the change in endorsement status with CMS's transition from the Physician Quality Reporting System (PQRS) to the Quality Payment Program (QPP).
0395 Paired Measure: Hepatitis C Ribonucleic Acid (RNA) Testing Before Initiating Treatment (paired with 0396)	Percentage of patients aged 18 years and older with a diagnosis of chronic hepatitis C who started antiviral treatment within the 12 month reporting period for whom quantitative hepatitis C virus (HCV) ribonucleic acid (RNA) testing was performed within 12 months prior to	In order to align the change in endorsement status with CMS's transition from the Physician Quality Reporting System (PQRS) to the Quality Payment Program (QPP).

Measure	Measure Description	Reason for Removal of Endorsement
	initiation of antiviral treatment	
0396 Paired Measure: Hepatitis C Virus (HCV) Genotype Testing Prior to Treatment (paired with 0395)	Percentage of patients aged 18 years and older with a diagnosis of chronic hepatitis C who started antiviral treatment within the 12 month reporting period for whom hepatitis C virus (HCV) genotype testing was performed within 12 months prior to initiation of antiviral treatment	In order to align the change in endorsement status with CMS's transition from the Physician Quality Reporting System (PQRS) to the Quality Payment Program (QPP).
0398 Hepatitis C: Hepatitis C Virus (HCV) Ribonucleic Acid (RNA) Testing Between 4-12 Weeks after Initiation of Treatment	Percentage of patients aged 18 years and older with a diagnosis of chronic hepatitis C who are receiving antiviral treatment for whom quantitative hepatitis C Virus (HCV) ribonucleic acid (RNA) testing was performed between 4-12 weeks after the initiation of antiviral treatment	In order to align the change in endorsement status with CMS's transition from the Physician Quality Reporting System (PQRS) to the Quality Payment Program (QPP).
0399 Paired Measure: Hepatitis C: Hepatitis A Vaccination (paired with 0400)	Percentage of patients aged 18 years and older with a diagnosis of hepatitis C who have received at least one injection of hepatitis A vaccine, or who have documented immunity to hepatitis A	In order to align the change in endorsement status with CMS's transition from the Physician Quality Reporting System (PQRS) to the Quality Payment Program (QPP).
0404 HIV/AIDS: CD4 Cell Count or Percentage Performed	Percentage of patients aged six months and older with a diagnosis of HIV/AIDS, with at least two CD4 cell counts or percentages performed during the measurement year at least 3 months apart	Developer is no longer able to support measure. Retired by Developer.
0408 HIV/AIDS: Tuberculosis (TB) Screening	Percentage of patients aged 3 months and older with a diagnosis of HIV/AIDS, for whom there was documentation that a tuberculosis (TB) screening test was performed and results interpreted (for tuberculin skin tests) at least once since the diagnosis of HIV infection.	Developer is no longer able to support measure. Retired by Developer.

Appendix A – NQF Member Voting Results

NQF MEMBER VOTING RESULTS

The nine recommended measures were approved with 67 % approval or higher. Representatives of three member organizations voted; no votes were received from Health Plan, Health Professional, Public/Community Health Agency or Supplier/Industry Councils. Results for each measure are provided below.

NQF Member Council	Voting Organizations	Eligible to Vote	Rate
Consumer	1	38	3%
Health Plan	0	21	0%
Health Professional	0	104	0%
Provider Organizations	1	110	1%
Public/Community Health Agency	0	15	0%
Purchaser	1	22	5%
QMRI	0	74	0%
Supplier/Industry	0	35	0%
All Councils	**3**	**419**	**1%**

2082 HIV Viral Load Suppression

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	0	1	100%
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	
All Councils	**3**	**0**	**0**	**3**	**100%**
Percentage of councils approving (>60%)					**100%**
Average council percentage approval					**100%**

*equation: Yes/ (Total - Abstain)

3210 HIV Viral Load Suppression – Legacy eMeasure

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	0	0	1	1	
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	
All Councils	**2**	**0**	**1**	**3**	**100%**
Percentage of councils approving (>60%)					**100%**
Average council percentage approval					**100%**

*equation: Yes/ (Total - Abstain)

2079 HIV Medical Visit Frequency

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	0	1	100%
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	
All Councils	**3**	**0**	**0**	**3**	**100%**
Percentage of councils approving (>60%)					**100%**
Average council percentage approval					**100%**

*equation: Yes/ (Total - Abstain)

3209 HIV Medical Visit Frequency – Legacy eMeasure

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	0	0	1	1	
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	

All Councils	2	0	1	3	100%
Percentage of councils approving (>60%)					100%
Average council percentage approval					100%

*equation: Yes/ (Total - Abstain)

2080 Gap in HIV Medical Visit

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	0	1	100%
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	
All Councils	**3**	**0**	**0**	**3**	**100%**
Percentage of councils approving (>60%)					**100%**
Average council percentage approval					**100%**

*equation: Yes/ (Total - Abstain)

2083 Prescription of HIV Antiretroviral Therapy

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	0	1	100%
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	
All Councils	**3**	**0**	**0**	**3**	**100%**
Percentage of councils approving (>60%)					**100%**
Average council percentage approval					**100%**

*equation: Yes/ (Total - Abstain)

3211 Prescription of HIV Antiretroviral Therapy – Legacy eMeasure

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	0	0	1	1	
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	
All Councils	**2**	**0**	**1**	**3**	**100%**
Percentage of councils approving (>60%)					**100%**
Average council percentage approval					**100%**

*equation: Yes/ (Total - Abstain)

0500 Severe Sepsis and Septic Shock: Management Bundle

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	0	1	100%
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	0	1	0	1	0%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	
All Councils	**2**	**1**	**0**	**3**	**67%**
Percentage of councils approving (>60%)					**67%**
Average council percentage approval					**67%**

*equation: Yes/ (Total - Abstain)

Voting Comments

Johns Hopkins Health Systems: We believe that measuring severe sepsis and septic shock is important to improve patient outcomes. It is explicitly stated that the components of the composite measure differ in the strength of the recommendation and the quality of the data, yet all have equal weight, in that any omission equals failure. Consider scoring each component of the measure to assist hospitals with knowing where to focus on improvement. For example, score 1 for delivering broad spectrum antibiotic coverage within 3 hours, 2 for delivery within 2 hours and 3 for delivery within 1 hour. Also it is essential to know that improvement in the bundle corresponds with improvement in outcomes. It is critical to measure that improvement in scores on the sepsis measure are accompanied by decreases in morbidity and mortality of sepsis.

3215 Adult Inpatient Risk Adjusted Sepsis Mortality

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	1	0	0	1	100%
Health Plan	0	0	0	0	
Health Professional	0	0	0	0	
Provider Organizations	1	0	0	1	100%
Public/Community Health Agency	0	0	0	0	
Purchaser	1	0	0	1	100%
QMRI	0	0	0	0	
Supplier/Industry	0	0	0	0	
All Councils	**3**	**0**	**0**	**3**	**100%**
Percentage of councils approving (>60%)					**100%**
Average council percentage approval					**100%**

*equation: Yes/ (Total - Abstain)

Appendix B – Measure Evaluation Summary Tables

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable; **Y**=Yes; **N**=No

2082 HIV Viral Load Suppression

Submission | Specifications

Description: Percentage of patients, regardless of age, with a diagnosis of HIV with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year. A medical visit is any visit in an outpatient/ambulatory care setting with a nurse practitioner, physician, and/or a physician assistant who provides comprehensive HIV care.

Numerator Statement: Number of patients in the denominator with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year

Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year

Exclusions: There are no patient exclusions.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Clinician Office/Clinic

Type of Measure: Outcome

Data Source: Laboratory, Paper Records

Measure Steward: Health Resources and Services Administration - HIV/AIDS Bureau

STANDING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-16; N-0**; 1b. Performance Gap: **H-10**; **M-3**; **L-0**; **I-0;**

Rationale:

- For the 2013 endorsement evaluation, the developer noted that viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. The developer also provided multiple guidelines for the administration of antiretroviral therapy and viral load monitoring intervals for adults, adolescents and pregnant women.
- For the current maintenance of endorsement evaluation, the Committee agreed with the developer that antiretroviral therapy and viral suppression reduce morbidity and mortality associated with HIV. The developer also submitted updated guidelines for the administration of antiretroviral therapy and viral load monitoring intervals for people living with HIV (PLWH).
- A Committee member pointed out CDC data that showed viral suppression rates were overestimated by 20% when looking at the last viral load test in a measurement period. Specifically, the analysis showed that not all patients that were suppressed at the end of the year were suppressed *throughout* the year. The developer clarified that this measure is not intended to be a durable suppression measure.
- The Committee reviewed 2010-2014 performance gap data from the Ryan White HIV/AIDS Program Services Report, with 65% viral load suppression among the 10th percentile of providers, and 94% among the 90th percentile.
- Overall, the Committee agreed that the measure met this criterion.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-12; M-4; L-0; I-0** 2b. Validity: **Accepted prior evaluation**

Rationale:

- For the 2013 endorsement evaluation, the developer conducted a signal-to-noise analysis using nine of 18 HIV Research Network (HIVRN) sites and 11,134 patients in 2010. The clinic-specific reliability ranged from 0.93 to 0.99 with a between clinic variance of 0.0066.
- For the current evaluation, the developer provided updated reliability testing using the beta binomial method to assess the signal-to-noise-ratio in over 800 clinics participating in the RSR database; median reliability ranged from 0.95 to 0.98 indicating high reliability.

2082 HIV Viral Load Suppression

- A Committee member noted that since reliability ranged from 0.29 to 0.98 that some clinics may not be performing well on the measure. The developer stated that lower signal-to-noise ratios generally came from clinics with small patient populations (i.e., <25 patients).
- A Committee member questioned why the measure is specified at less than 200 copies/mL. The developer cited HHS guidelines, which indicate that although tests can quantify a viral load down to individual copies of the virus, it is generally accepted that 200 is the threshold for viral load suppression.
- Another Committee member asked the developer to define "comprehensive HIV care." The developer noted that HIV care is in various models, but ultimately they define comprehensive care as being provided by someone who is addressing the patient's HIV care (i.e., assessing whether the client is on antiretroviral medications and that viral load testing has been performed).
- One Committee member asked how the measure accounts for providers who could game the measure by choosing not to see or to create an unfavorable environment for non-adherent patients. The developer noted that it would be difficult to capture that information at the individual provider level since providers do not have a lot of control over the patients they see. The developer also noted that there is a measure (#2079) that assesses whether patients come back to the same clinic.
- In discussing what constitutes a medical visit, the Committee questioned whether non-face to face visits are included in the measure. The developer clarified that face-to-face, video visits, or another type of visit that is documented by the provider as a medical visit. If a patient comes in for lab work, to meet with an ancillary staff member or to pick up paperwork, the visit is not counted as a medical visit. The measure does not address whether visits are billable. The developer also clarified that patients are counted in the denominator if they came in for at least one visit.
- The Committee noted that the developer used a technical group and Ryan White grant recipients to test the face validity of the measure. A Committee member questioned whether the same group of people established the measure and assessed face validity. The developer clarified that these were the same group of people.
- A Committee member expressed concern that although the measure is not risk adjusted, since people living with HIV (PLWH) are a marginalized population, that the measure could be used in MIPS, a pay for performance program. The Committee member expressed concern that a provider could be penalized based on poor performance on this measure if they provided medical care to populations who are non-adherent to medical care (e.g., people experiencing homelessness, active substance users).
- The developer stated that existing methods for risk adjustment do not apply to this measure because it is currently not used in a payment program. However, the eMeasure version of this measure was submitted for consideration in the Centers for Medicare & Medicaid Services (CMS) Merit-based Incentive Payment System (CMS MIPS); therefore, the developer will work with CMS to develop appropriate methods for risk adjustment of the eMeasure.
- The Committee emphasized that there are concerns among PLWH that as measurement of HIV care is increasingly implemented, that providers will make an unfavorable environment for PLWH.
- The Committee questioned how the measure accounts for patients who have access to the care but choose not to receive it. The developer clarified that the intent of the measure is not to achieve 100%, nor is it a measure of the patient's compliance or ability to participate in care. The measure is intended to assess the clinic's ability to support a client in reaching viral load suppression.
- Upon a vote, the Committee agreed the measure met the reliability criteria.
- For the 2013 endorsement evaluation, face validity was established through a technical work group established for the development of the measures. This measure was found to be important, usable, and feasible by the technical work group overseeing the development of this measure and several others.
- For the current evaluation, the developer did not provide updated validity testing because they intend to replace this paper-based measure with #3210, the electronically specified version of this measure. The Committee agreed that the previous validity testing results were sufficient and accepted the prior evaluation without further discussion and vote.

3. Feasibility: H-14; M-1; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

2082 HIV Viral Load Suppression

- All data elements for this measure are generated during routine provision of care and are in defined fields in the electronic health record. Without further discussion, the Committee agreed the measure met this criterion.

4. Usability and Use: H-15; M-0; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The measure is publicly reported and used by accountability programs including MIPS and PQRS. The measure has shown improvement in viral load suppression and no potential harms were identified in measure implementation. Without further discussion, the Committee agreed the measure met this criterion.

5. Related and Competing Measures

- This measure is related to the following:
  - 0407 HIV/AIDS: HIV RNA Control After Six Months of Potent Antiretroviral Therapy
  - 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis
  - 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis
  - 2079 HIV Medical Visit Frequency
  - 2080 Gap in HIV Medical Visits
  - 2083 Prescription of HIV Antiretroviral Therapy
  - 3211 Prescription of HIV Antiretroviral Therapy
  - 3210 HIV Viral Suppression
  - 3209 HIV Medical Visit Frequency

Per the developer, this measure is harmonized with all measures except for #0405 and #0409; there are plans to harmonize with #0405 and #0409.

Standing Committee Recommendation for Endorsement: **Y-16; N-0**

6. Public and Member Comment

- One comment was submitted in support of the Committee's recommendation for endorsement.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3210 HIV Viral Load Suppression

Submission | Specifications

Description: Percentage of patients, regardless of age, with a diagnosis of HIV with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year.

Numerator Statement: Patients with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year. The outcome being measured is HIV viral suppression.

Denominator Statement: Patients, regardless of age, diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year who had at least one medical visit in the measurement year. The target population for this measure is all people living with HIV.

Exclusions: There are no patient exclusions.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Clinician Office/Clinic

Type of Measure: Outcome,

Data Source: Electronic Health Record (Only), Other

Measure Steward: Health Resources and Services Administration - HIV/AIDS Bureau

STANDING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Applied the vote from #2082**; 1b. Performance Gap: **Applied the vote from #2082**

Rationale:

- The Committee acknowledged that this measure shares the same evidence as #2082. Because it is a legacy eMeasure, there are no performance data. The Committee agreed to apply the vote from #2082 to this criterion.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: H-0; **M-15; L-1; I-0** 2b. Validity: **H-0; M-16; L-0; I-0**

Rationale:

- This measure was tested using 34 synthetic cases with the Bonnie tool. Results showed 100% coverage and all 34 cases passed the measure as expected, demonstrating that the measure logic works as constructed. All test cases with missing data performed according to the HQMF standard as expected. Testing also included specific ways to search for patients that might fall at the edge of the measure specifications.
- Without further discussion, the Committee agreed the measure met these criteria.

3. Feasibility: H-12; M-4; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- Feasibility testing showed that the measure is feasible but the Committee questioned why the variable "Encounter Performed: Face-to-Face Interaction" and "Patient Characteristic Payer" scored a 2 out of 3 on the feasibility scorecard. The developer clarified that "Face-to-Face Interaction" scored lower because the value set is defined in SNOMED whereas the other encounter value sets are defined in CPT. The developer further clarified that the variable "Patient Characteristic Payer" is a supplemental data element required to be submitted for measures in federal programs but is not used in the measure logic.
- The Committee asked the developer to explain why they expect feasibility to improve from 98.89% to 99.44%. The developer clarified that the addition of the SNOMED codes will increase the feasibility.
- Overall, the Committee agreed the measure met this criterion.

4. Usability and Use: H-8; M-8; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

3210 HIV Viral Load Suppression

Rationale:

- The Committee noted that the measure was reviewed by NQF's Measure Application Partnership (MAP) for consideration in CMS' MIPS program. MAP recommended support of this measure for rulemaking with the condition that the Infectious Disease Standing Committee review the performance data to ensure a gap in care continues to exist.
- The developer clarified that this measure is intended to be used in an accountability program. Without further discussion, the Committee agreed the measure met this criterion.

5. Related and Competing Measures

- This measure is related to the following:
  - 0407 HIV/AIDS: HIV RNA Control After Six Months of Potent Antiretroviral Therapy
  - 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis
  - 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis
  - 2079 HIV Medical Visit Frequency
  - 2080 Gap in HIV Medical Visits
  - 2083 Prescription of HIV Antiretroviral Therapy
  - 3211 Prescription of HIV Antiretroviral Therapy
  - 3210 HIV Viral Suppression
  - 3209 HIV Medical Visit Frequency
- Per the developer, this measure is harmonized with all measures except for #0405 and #0409; there are plans to harmonize with #0405 and #0409.

Standing Committee Recommendation for Endorsement: **Y-16; N-0**

6. Public and Member Comment

- One comment was submitted in support of the Committee's recommendation for endorsement.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2079 HIV Medical Visit Frequency

Submission | Specifications

Description: Percentage of patients, regardless of age, with a diagnosis of HIV who had at least one medical visit in each 6-month period of the 24-month measurement period with a minimum of 60 days between medical visits. A medical visit is any visit in an outpatient/ambulatory care setting with a nurse practitioner, physician, and/or a physician assistant who provides comprehensive HIV care.

Numerator Statement: Number of patients in the denominator who had at least one medical visit in each 6-month period of the 24-month measurement period with a minimum of 60 days between first medical visit in the prior 6-month period and the last medical visit in the subsequent 6-month period. (Measurement period is a consecutive 24-month period of time.)

Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the first 6 months of the 24-month measurement period.

Exclusions: Patients who died at any time during the 24-month measurement period.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Clinician Office/Clinic

Type of Measure: Process

Data Source: Paper Records

Measure Steward: Health Resources and Services Administration - HIV/AIDS Bureau

STANDING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **H-0; M-2; L-1; I-13**; Evidence with Exception: **Y-14; N-2**; 1b. Performance Gap: **H-13; M-3; L-0; I-0**

Rationale:

- This measure calculates the percent of clients retained in care over a 24-month time period. Patients are counted in the numerator if they have had at least one medical visit at least 60 days apart in each 6 months of the measurement period.
- The Committee acknowledged the developer submitted updated evidence for the systematic monitoring of retention in care they could include visit adherence, gaps in care, and number of visits during a specified visit. A Committee member questioned whether there was evidence to support the 6-month interval and the 60-days between medical visits, especially when the guidelines state that clinically stable clients can be seen less frequently.
- The developer stated that the 6-month time period was selected as a 'middle ground' based on DHHS guidelines that outline the frequency of labs performed and medical visits. The 60-day period between medical visits was selected so as to not count patients, who for instance, have medical visits on consecutive days.
- Another Committee member expressed concern for programs receiving Ryan White program funds who must report this measure. The committee member noted that some programs may be forced to see stable clients more frequently just to meet the measure's requirements.
- The Committee then debated that the measure should focus on frequency of viral load testing for stable clients who only need to see a provider for lab work. They also questioned how the measure accounts for not penalizing providers who have durably suppressed clients. Some Committee members supported that the measure focus was acceptable since retention in care has been associated with improved clinical outcomes. A Committee member commented that medical visit frequency is more controlled by the health system than by the individual provider.
- The developer further clarified that they do not expect to reach 100% adherence on this measure and the measure is applicable broadly across all types of providers caring for any patient living with HIV.
- The Committee was conflicted as to whether the measure should be used for public health rather than quality improvement purposes.
- Upon a vote, the Committee did not pass the measure on the Evidence criterion but agreed to vote on whether the measure warranted an exception to the evidence. The Committee discussed the importance of the measure on outcomes for people living with HIV but did not have evidence that explicitly stated that the measure as specified would lead to those outcomes. After a full discussion,

the Committee opted to invoke the exception to the evidence criterion and agreed that medical visit frequency needed to be assessed since patients cannot be treated if they have not been seen by a provider. Ultimately, a majority of the Committee agreed that providers should be held accountable for this measure in the absence of empiric evidence and passed the measure on this criterion.

- The Committee noted that medical visit frequency had increased from 67% to 73%, with disparities – similar to those seen in #2082 – among race/ethnic, gender, and age groups.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-13; M-2; L-0; I-0** 2b. Validity: **Accepted previous Evaluation**

<u>Rationale</u>:

- The Committee reviewed testing data from the RSR (covering 11 sites and more than 17,000 patients) that showed a median reliability of 0.97; upon a vote, the Committee agreed the measure met this criterion.
- Committee members questioned whether the measure is stratified to determine the model of care the patient receives (i.e., care from a primary care provider or an infectious disease provider). The developer clarified that the measure is not stratified by the type of provider.
- As was discussed for #2082, the developer clarified that tele-visits or other advanced methods count as a medical visit; the measure does not specify how the visit is delivered or whether the visit is billable.
- In response to the Committee's question as to why validity testing was not updated, the developer responded they used their resources for the testing and development of the eMeasures. The Committee then noted the performance data demonstrated the measure was able to identify differences in performance among providers. The Committee chose to accept the previous evaluation on this criterion.

3. Feasibility: H-13; M-1; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

<u>Rationale</u>:

- The Committee agreed that the measure was feasible with all data elements collected and generated as part of routine delivery of care. Without further discussion, the Committee agreed the measure met this criterion.

4. Usability and Use: H-11; M-4; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

<u>Rationale</u>:

- The measure is currently in use in CMS' Physician Quality Report System (PQRS), Value Based Payment Modifier (VBPM), and MIPS programs. Without further discussion, the Committee agreed the measure met this criterion.

5. Related and Competing Measures

- This measure is related to the following:
  - 2080 Gap in HIV Medical Visits
  - 2082 HIV Viral Suppression
  - 2083 Prescription of HIV Antiretroviral Therapy
  - 3211 Prescription of HIV Antiretroviral Therapy
  - 3210 HIV Viral Suppression
  - 3209 HIV Medical Visit Frequency
  - 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis
  - 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis
- Per the developer, the measure is harmonized with the first six measures listed above. For these six measures the target population is the same (i.e., people living with HIV), however the measure focus is different (gaps in visit, prescription of ARV therapy, and viral load suppression).

2079 HIV Medical Visit Frequency
• NQF #0405 and #0409 have been deferred for maintenance of endorsement. There are no additional steps the developer must take since the measure focus is different (HIV patients receiving PCP prophylaxis and those screened for STDs).
Standing Committee Recommendation for Endorsement: **Y-15; N-0**
6. Public and Member Comment • One comment was submitted in support of the Committee's recommendation for endorsement.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

3209 HIV Medical Visit Frequency

Submission | Specifications

Description: Percentage of patients, regardless of age, with a diagnosis of HIV who had at least one medical visit in each 6-month period within 24 months with a minimum of 60 days between medical visits. A medical visit is any visit in an outpatient/ambulatory care setting with a nurse practitioner, physician, and/or a physician assistant who provides comprehensive HIV care.

Numerator Statement: Patients who had at least one medical visit in each 6-month of a consecutive 24 month period with a minimum of 60 days between first medical visit in the prior 6-month period and the last medical visit in the subsequent 6-month period.

Denominator Statement: Patients, regardless of age, diagnosed with HIV during the first 3 months of the year preceding the measurement period or prior to the measurement period with at least one medical visit in the first 6 months of the year preceding the measurement period. The target population for this measure is all people living with HIV.

Exclusions: Patients who died at any time during the measurement period or the 12 months preceding the measurement period.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Clinician Office/Clinic

Type of Measure: Process

Data Source: Electronic Health Record (Only)

Measure Steward: Health Resources and Services Administration - HIV/AIDS Bureau

STANDING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Applied the vote from #2079**; 1b. Performance Gap: **Applied the vote from #2079**

Rationale:

- The Committee acknowledged that this measure shares the same evidence as #2079. Because it is a legacy eMeasure, there is no performance data. The Committee agreed to apply the vote from #2079 to this criterion.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-0**; **M-10; L-5; I-0** 2b. Validity: **H-0**; **M-14; L-1; I-0**

Rationale:

- This measure was tested using 64 synthetic cases with the Bonnie tool. Results showed 100% coverage and all 64 cases passed the measure as expected, demonstrating that the measure logic works as constructed. Testing also included specific ways to search for patients that might fall at the edge of the measure specifications.
- The Committee noted that the measure excludes patients who die during the measurement period. Another Committee member stated that the measure should exclude patients who are incarcerated during the measurement period since providers could be penalized on this measure if they provide care to specific populations that experience incarceration. The developer clarified that there is no standardized variable for incarceration in electronic health records.
- Without further discussion, the Committee agreed this measure met this criterion.

3. Feasibility: H-9; M-6; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- Feasibility testing showed that the measure is feasible at 98.21% and will increase to 98.81%.
- The Committee questioned why the variable “Encounter Performed: Face-to-Face Interaction” and “Patient Characteristic Payer” scored a 2 out of 3 on the feasibility scorecard. The developer clarified in the discussion of #3210 that “Face-to-Face Interaction” scored lower because the value set is defined in SNOMED whereas the other encounter value sets are defined in CPT. The “Patient Characteristic

3209 HIV Medical Visit Frequency
Payer" variable is a supplemental data element required to be submitted for measures in federal programs but is not used in the measure logic. • Overall, the Committee agreed the measure met this criterion.
4. Usability and Use: H-7; M-8; L-0; I-1 *(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)* Rationale: • The Committee noted that the measure is not publically reported. The measure is planned to be used in MIPS. Without further discussion, the Committee agreed the measure met this criterion.
5. Related and Competing Measures • This measure is related to the following: o 2080 Gap in HIV Medical Visits o 2082 HIV Viral Suppression o 2083 Prescription of HIV Antiretroviral Therapy o 3211 Prescription of HIV Antiretroviral Therapy o 3210 HIV Viral Suppression o 3209 HIV Medical Visit Frequency o 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis o 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis • Per the developer, the measure is harmonized with the first six measures listed above. For these six measures the target population is the same (i.e., people living with HIV), however the measure focus is different (gaps in visit, prescription of ARV therapy, and viral load suppression). • NQF #0405 and #0409 have been deferred for maintenance of endorsement. There are no additional steps the developer must take since the measure focus is different (HIV patients receiving PCP prophylaxis and those screened for STDs).
Standing Committee Recommendation for Endorsement: **Y-16; N-0**
6. Public and Member Comment • One comment was submitted in support of the Committee's recommendation for endorsement.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

2080 Gap in HIV Medical Visits

Submission | Specifications

Description: Percentage of patients, regardless of age, with a diagnosis of HIV who did not have a medical visit in the last 6 months of the measurement year. A medical visit is any visit in an outpatient/ambulatory care setting with a nurse practitioner, physician, and/or a physician assistant who provides comprehensive HIV care.

Numerator Statement: Number of patients in the denominator who did not have a medical visit in the last 6 months of the measurement year (Measurement year is a consecutive 12-month period of time).

Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV who had at least one medical visit in the first 6 months of the measurement year. (The measurement year can be any consecutive 12-month period.)

Exclusions: Patients who died at any time during the measurement year.

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Clinician : Group/Practice, Facility

Setting of Care: Clinician Office/Clinic

Type of Measure: Process

Data Source: Other, Paper Records

Measure Steward: Health Resources and Services Administration-HIV/AIDS Bureau

STANDING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **H-0**; **M-2**; **L-0**; **I-11** 1b. Performance Gap: **H-3**; **M-10**; **L-0**; **I-0;** Evidence Exception: **Y-12; N-1**

Rationale:

- The Committee questioned the value of the measure over the related measure, #2079. The developer clarified that this measure looks at a shorter time period and that some users pair this measure with the longer term retention measure (#2079).
- A Committee member acknowledged the importance of retention in care, but did not believe there was evidence related to two medical visits and health outcomes. The Committee asked the developer to clarify the definition of a medical visit. The developer noted they do not specify the mode of the visit (face-to-face vs. tele-health) but that tele-health visits are not ruled out.
- In response to the Committee's question as to how a gap in medical care is defined, the developer stated that if a patient has a medical visit in the first six months of the measurement period but not within the second six months then the client would have experienced a gap in care. Based on this definition, their data indicates that 21% of patients had a gap in medical visits. Better performance for this measure is indicated by a lower rate e.
- A Committee member noted that the guidelines have changed to say that viral load should be assessed every six months but that durably suppressed clients can be seen less frequently.
- The Committee agreed that the evidence provided was insufficient; however, they agreed retention in care is important and voted to pass the measure on evidence with an exception. Notably, the Committee agreed that the measure could be used to help providers prioritize clients who have not been in care.
- The Committee noted that the performance gap had increased from 2010 to 2014 (i.e., more people are not getting regular care). Without further discussion, the Committee agreed the measure met this criterion.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Accepted prior evaluation** 2b. Validity: **Accepted prior evaluation**

Rationale:

- The Committee reviewed the testing data from over 800 participants in the Ryan White HIV/AIDS Program (RWHAP) that showed a median reliability of 0.973; without further discussion agreed to accept the prior evaluation of this measure.

2080 Gap in HIV Medical Visits

- In discussion of measure validity, the Committee noted that new testing was provided but that face validity was completed. They also noted the exclusion for patients who died (<1%) had minimal impact on the overall score.
- The developer clarified they are field-testing an electronic version of this measure. Without further discussion, the Committee agreed to accept the prior evaluation of this measure.

3. Feasibility: H-10; M-3; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee acknowledged that data are generated or collected by and used by health care professionals during the provision of care. All data elements are in defined fields in electronic health records. There are no fees, licenses, or other requirements to use this measure.
- A Committee member commented that the measure is calculated in the inverse and it could be confusing to those who use the measure.

4. Usability and Use: H-10; M-3; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The Committee asked the developer to explain the measure selection process for recipients of RWHAP grant funds. The developer clarified that RWHAP does not specify which measures recipients must use.
- The Committee discussed that the measure is the most useful to determine how to schedule patients, with Committee members anecdotally sharing they find the measure valuable.
- A Committee member questioned how the measure is implemented in paper records. Other Committee members responded that smaller providers with paper records have to manually count the measure and that larger centers with paper records may find the manual implementation of the measure to be difficult.

5. Related and Competing Measures

- This measure is related to the following:
  - 2079 HIV Medical Visit Frequency
  - 2082 HIV Viral Suppression
  - 2083 Prescription of HIV Antiretroviral Therapy
  - 3211 Prescription of HIV Antiretroviral Therapy
  - 3210 HIV Viral Suppression
  - 3209 HIV Medical Visit Frequency
  - 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis
  - 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis
- Per the developer, the measure is harmonized with the first six measures listed above. For these six measures the target population is the same (i.e., people living with HIV), however the measure focus is different (medical visit frequency, prescription of ARV therapy, and viral load suppression).
- NQF #0405 and #0409 have been deferred for maintenance of endorsement. There are no additional steps the developer must take since the measure focus is different (HIV patients receiving PCP prophylaxis and those screened for STDs).

Standing Committee Recommendation for Endorsement: **Y-13**; **N-0**

6. Public and Member Comment

- One comment was submitted in support of the Committee's recommendation for endorsement.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

2083 Prescription of HIV Antiretroviral Therapy

Submission | Specifications

Description: Percentage of patients, regardless of age, with a diagnosis of HIV prescribed antiretroviral therapy for the treatment of HIV infection during the measurement year. A medical visit is any visit in an outpatient/ambulatory care setting with a nurse practitioner, physician, and/or a physician assistant who provides comprehensive HIV care.

Numerator Statement: Number of patients from the denominator prescribed HIV antiretroviral therapy during the measurement year.

Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year.

Exclusions: There are no patient exclusions.

Adjustment/Stratification:

Level of Analysis: Facility

Setting of Care: Clinician Office/Clinic

Type of Measure: Process

Data Source: Other, Paper Records, Pharmacy

Measure Steward: Health Resources and Services Administration - HIV/AIDS Bureau

STANDING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Accepted previous evaluation**; 1b. Performance Gap: **H-12**; **M-0**; **L-0**; **I-0;**

Rationale:

- A Committee member noted that the evidence base for this measure addresses viral suppression but not the prescription of antiretroviral (ARV) therapy, which is the measure's focus. The Committee member pointed out that they should consider whether the mere *prescription* of antiretroviral therapy rather than the *receipt* of the therapy by the patient is adequate evidence.
- The developer noted that there is not a way to measure whether patients are picking up their medications from the pharmacy and that this measure is used in tandem with #2082.
- The Committee acknowledged that the measure is based on a strong recommendation for ARV (i.e., that the patient is taking the medication) but noted there is a difference between writing a prescription (i.e., the focus of this measure) and a patient actually receiving the medication. The Committee stated that although the measure is valuable, there does not seem to be a link between *prescribing* ARV and viral suppression.
- Ultimately, the Committee decided not to re-vote on the evidence and accepted the previous evaluation.
- From 2010 to 2014, measure performance improved from 68.4% to 77.6%, with a median score of 90% in 2014. The Committee reviewed data that showed disparities among age, gender, and racial/ethnic groups. Overall, the Committee agreed the measure met this criterion.

2. Scientific Acceptability of Measure Properties: The measure does meet the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Accepted prior evaluation** 2b. Validity: **Accepted prior evaluation**

Rationale:

- The Committee reviewed testing data from the Ryan White HIV/AIDS Program Services Report (RSR) (covering 2000 RWHAP recipients) showing reliability at 0.99 and found the testing method appropriate; without further discussion the Committee accepted the previous evaluation on this criterion.
- The Committee noted that the measure is abstracted from paper and electronic records and questioned why the measure does not exclude patient death. The developer sited that some providers are starting to use electronic health records and noted that death was not an exclusion for this measure because HIV mortality is extremely low and the increased burden of data collection and analysis does not add value to the measure.
- The Committee accepted the previous evaluation for reliability and validity.

2083 Prescription of HIV Antiretroviral Therapy

3. Feasibility: H-11; M-1; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committed note the data elements are in electronic records, are in use and are readily available. Without further discussion, the Committee agreed the measure met this criterion.

4. Usability and Use: H-9; M-3; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale

- The Committee acknowledges that the measure is in use and is publically reported in PQRS and MIPS. The Department of Health and Human Services selected this measure as a core HIV indicator. The Committee then agreed that the measure met this criterion.

5. Related and Competing Measures

- This measure is related to the following:
  - 2079 HIV Medical Visit Frequency
  - 2080 Gap in HIV Medical Visits
  - 2082 HIV Viral Suppression
  - 3211 Prescription of HIV Antiretroviral Therapy
  - 3210 HIV Viral Suppression
  - 3209 HIV Medical Visit Frequency
  - 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis
  - 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis
- Per the developer, the measure is harmonized with the first six measures listed above. For these six measures the target population is the same (i.e., people living with HIV), however the measure focus is different (gaps in visits, medical visit frequency, and viral load suppression).
- NQF #0405 and #0409 have been deferred for maintenance of endorsement. There are no additional steps the developer must take since the measure focus is different (HIV patients receiving PCP prophylaxis and those screened for STDs).

Standing Committee Recommendation for Endorsement: **Y-12**; **N-0**

6. Public and Member Comment

- One comment was submitted in support of the Committee's recommendation for endorsement.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

3211 Prescription of HIV Antiretroviral Therapy

Submission | Specifications

Description: Percentage of patients, regardless of age, with a diagnosis of HIV prescribed antiretroviral therapy for the treatment of HIV infection during the measurement year. A medical visit is any visit in an outpatient/ambulatory care setting with a nurse practitioner, physician, and/or a physician assistant who provides comprehensive HIV care.

Numerator Statement: Number of patients from the denominator prescribed HIV antiretroviral therapy during the measurement year.

Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year.

Exclusions: There are no patient exclusions.

Adjustment/Stratification: No risk adjustment or risk stratification

3211 Prescription of HIV Antiretroviral Therapy

Level of Analysis: Facility

Setting of Care: Clinician Office/Clinic

Type of Measure: Process

Data Source: Electronic Health Record (Only)

Measure Steward: Health Resources and Services Administration - HIV/AIDS Bureau

STANDING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Applied the vote from #2083**; 1b. Performance Gap: **Applied the vote from #2083**

<u>Rationale</u>:

- The Committee acknowledged that this measure shares the same evidence as #2083. Because it is a legacy eMeasure, there is no performance data. The Committee agreed to apply the vote from #2083 to this criterion.

2. Scientific Acceptability of Measure Properties: <u>The measure does meet the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-0; M-12; L-0; I-0** 2b. Validity: **H-0; M-12; L-0; I-0**

<u>Rationale</u>:

- This measure was tested using 34 synthetic cases with the Bonnie tool. Results showed 100% coverage and all 34 cases passed the measure as expected, demonstrating that the measure logic works as constructed. Testing also included specific ways to search for patients that might fall at the edge of the measure specifications.
- The Committee agreed that the measure specifications are consistent with the evidence and noted that a panel of experts looked at each synthetic cases and assigned an outcome which then correlated with the results of the Bonnie testing, which demonstrates that the measure logic works.

3. Feasibility: H-9; M-3; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

<u>Rationale</u>:

- The Committed acknowledge measure feasibility at 98.33% and noted that data elements are in electronic records, are in use and are readily available. The Committee agreed the measure met this criterion.

4. Usability and Use: H-7; M-5; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

<u>Rationale</u>

- The Committee acknowledges that the measure usability is similar to #2083 and without further discussion agreed the measure met this criterion.

5. Related and Competing Measures

- This measure is related to the following:
  - 2079 HIV Medical Visit Frequency
  - 2080 Gap in HIV Medical Visits
  - 2082 HIV Viral Suppression
  - 3210 HIV Viral Suppression
  - 3209 HIV Medical Visit Frequency
  - 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis
  - 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis
- Per the developer, the measure is harmonized with the first six measures listed above. For these six measures the target population is the same (i.e., people living with HIV), however the measure focus is different (gaps in visits, medical visit frequency, and viral load suppression).

3211 Prescription of HIV Antiretroviral Therapy
• NQF #0405 and #0409 have been deferred for maintenance of endorsement. There are no additional steps the developer must take since the measure focus is different (HIV patients receiving PCP prophylaxis and those screened for STDs).
Standing Committee Recommendation for Endorsement: **Y-12**; **N-0**
6. Public and Member Comment • One comment was submitted in support of the Committee's recommendation for endorsement.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

3215 Adult Inpatient Risk Adjusted Sepsis Mortality

Submission | **Specifications**

Description: Annual risk adjusted inpatient mortality rate for adult patients (aged 18 and over) admitted to acute care hospitals with diagnosis of severe sepsis or septic shock. The measure includes patients in acute care hospital settings over one year timeframe who had, either on admission, or during their hospital stay, a clinical diagnosis of severe sepsis (now referred to as 'sepsis') or septic shock using criteria described in the International Sepsis Definitions (Sepsis-2)

Hospitals were required to submit a protocol for early identification and treatment of severe sepsis or septic shock. Subsequent to protocol submission, hospitals were required to submit 100% of their patient cases to a data collection portal using a standardized data dictionary (see relevant sections for details). Numerous data elements including patient demographics and comorbidities among other patient care details were reported. A random sample of the data submissions were validated for accuracy. The full adult data for discharges within calendar year 2015 was used to generate statewide and hospital-specific risk adjusted mortality rates for the calendar year.

Numerator Statement: Outcome is risk adjusted inpatient mortality rate for adult patients (18 and over) admitted to an acute care hospital with a diagnosis of severe sepsis or septic shock or who develop severe sepsis or septic shock during their hospital stay.

Denominator Statement: All adult patient discharges (18 and over) in a calendar year with a diagnosis of severe sepsis or septic shock on admission or at any time during their hospital stay. This may include multiple admissions of the same patient during the measurement year. Denominator includes all cases identified using any means (administrative, registry, electronic health records, billing data, etc.), either prospectively, retrospectively, or both, that meet the International consensus definition (Sepsis- 2) of severe sepsis or septic shock.

Exclusions: Patients with advanced directives in place prior to episode of sepsis which specifically restrict any hospital specific sepsis protocol interventions or who decline (or their proxy declines) treatment for sepsis. Patients who have been transferred from one acute care hospital to another are excluded.

Adjustment/Stratification: Multivariate logistic regression model

Level of Analysis: Facility

Setting of Care: Hospital : Acute Care Facility

Type of Measure: Outcome

Data Source: Claims (Other), EHRs Hybrid, Laboratory, Management Data, Non-Medical Data, Paper Records, Pharmacy, Registry

Measure Steward: New York State Department of Health, Office of Quality and Patient Safety

STEERING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-16**; **N-0**; 1b. Performance Gap: **H-12**; **M-3**; **L-0**; **I-1**

Rationale:

- Mortality is an important outcome for patients with sepsis and according to the developer, mortality rates are high and show significant variability across acute care hospitals unrelated to patient factors.
- The developer suggested that hospitals are able to influence mortality rates using early sepsis detection approaches coupled with rapid delivery of basic resuscitation interventions including the use of adequate intravenous fluids, antibiotics, blood pressure support medications and dynamic clinical monitoring for response.
- The Committee agreed that the developer clearly identified how healthcare facilities and providers influence sepsis mortality outcomes.
- The developer provided the risk-adjusted probability of inpatient sepsis mortality rates from 179 hospitals and 43,204 patients in New York State from January 1, 2015 – December 31, 2015 for this newly developed measure. The mean performance rate in Q1 2015 was 30.4%, 28.9% in Q2 2015, 28.8% in Q3 2015 and 28.4% in Q5 2015. The performance rates in 2015 ranged from a minimum of 1.0% (Q1-Q4) to a maximum of 95.0% (Q4). The developer also provided the probability of inpatient sepsis mortality rates by population group, which included race/ethnicity, gender, age and insurance/payer. White, non-Hispanics had a rate of 28.2%; Black, non-Hispanics had a rate of 31.5%;

and Hispanics had a rate of 26.3%. Rates based on gender were similar with 28.4% for females and 29.8% for males. Patients 70 to 80+ years old had rates from 31.0% to 33.9%. The probability of inpatient sepsis mortality also varied based on insurance/payer. Patients with Medicare had a rate of 30.6%; Medicaid patients had a rate of 26.2%; Private pay and/or HMO patients had a rate of 27.1%; self-pay patients had the highest probability of inpatient sepsis morality, 34.3%.

- The Committee recognized the variability in coding practices related to sepsis and the potential impact on the denominator of this measure. After a lengthy discussion, the Committee agreed that the data presented by the developer demonstrated significant variation and an opportunity for improvement in inpatient sepsis mortality across hospitals.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-2**; **M-8**; **L-0**; **I-6** 2b. Validity: **H-4**; **M-8**; **L-1**; **I-3**

<u>Rationale</u>:

- The developer conducted data element validity testing, which counted for data element reliability testing as well.
- The dataset included hospitals in New York State that were required to develop and implement early recognition and treatment protocols for sepsis. As part of this statewide initiative, hospitals were required to submit quarterly clinical data to the New York State Department of Health to be evaluated for protocol use, adherence to time interventions and patient outcomes, including mortality.
    - The dataset included 179 hospitals with 43,204 patients diagnosed with severe sepsis and septic shock from January 1, 2015 to December 31, 2015.
    - The dataset used to develop the logistic regression model included:
        - Development sample: 38,884 (90%) patients; 179 hospitals
        - Validation sample: 4,319 (10%) patients; 160 hospitals
- The developer validated the accuracy of the data submission from the hospitals against manual chart abstraction by external auditors (Audit Results), which is considered the gold standard. The developer calculated the percent agreement between the hospital submissions to the chart-abstracted data. Percent agreement from the audit results ranged from 89.9% to 99.1% for the following data elements:
    - Site of infection: 98.9%
    - Lower Respiratory Infection: 98.8%
    - Mechanical Ventilation: 97.8%
    - Age (Date of Birth): 99.2%
    - Thrombocytopenia: 97.7%
    - Septic Shock: 98.4%
    - Serum Lactate (Lactate Level): 93.9%
    - Metastatic Cancer: 97.1%
    - Lymphoma, Leukemia, Multiple Myeloma: 99.1%
    - Square Root of Comorbidity Count (Range of Comorbidities): 89.9% - 99.1%
- The developer noted that the data elements race, ethnicity, payer and admission source were not manually audited but were aligned to state administrative datasets to ensure accuracy.
- The developer did not provide sensitivity, specificity, positive predictive value (PPV) and negative predictive values (NPV) in addition to percent agreement.
- The developer clarified that if a case was initially coded as severe sepsis or septic shock but the manual review found that the case did not meet the clinical definition of severe sepsis or septic shock, the hospital was able to exclude the case.
- A Committee member asked if the developer had compared patients that present to the emergency department (ED) with sepsis vs. patients that develop sepsis while hospitalized. The developer replied that they intend to continue researching the differences in outcomes for patients that present to the ED with sepsis vs. patients that acquire sepsis in the hospital.
- The Committee did not express any other concerns with the reliability of the measure and agreed it met the reliability criterion.
- Empirical validity testing of the measure score was assessed by comparing the performance of the risk-adjusted model in the development sample to the validation sample.

  - The developer used the Hosmer-Lemeshow goodness of fit test to assess the observed and expected mortality rates in the development and validation samples.
    - The development dataset was split into group sizes of 10, 100, 500 and 1,000. The p-values were 0.568, 0.972, 0.735, and 0.735, respectively.
    - The validation dataset was split into group sizes of 10, 50, 100 and 150. The p-values were 0.651, 0.977, 0.985, and 0.974, respectively.
- The performance of the risk-adjustment model was similar in the development and validation datasets. The areas under the receiver operating characteristic (ROC) curve (or *c*-statistic) were **0.770** and **0.773**, respectively. A *c*-statistic is a model of discrimination statistic. A *c*-statistic of 0.77 means that 77.0% of all possible pairs of patients – one who died and one who lived – the model correctly assigned a higher probability to those who died. The similar *c*-statistics indicates good model discrimination.
- A Committee member questioned whether a *c*-statistic of 0.7 was sufficient. NQF staff responded that although NQF staff does not have specific statistical thresholds, generally, a *c*-statistic of at least 0.70 is considered acceptable.
- Exclusions include patients with advanced directives in place prior to episode of sepsis which specifically restrict any hospital specific sepsis protocol interventions or who decline (or their proxy declines) treatment for sepsis and patients who have been transferred from one acute care hospital to another. The developer maintained that keeping patients in the dataset that had an advanced directive or declined intervention would bias the results towards higher hospital mortality.
- In the pre-evaluation comments, one of the Committee members questioned why excluding patients with multiple admissions is appropriate because this would artificially increase the mortality rate; however, patients with multiple admissions are not excluded from the denominator. The denominator details state that multiple admissions of the same patient during the measurement year are included.
- This measure is risk-adjusted using a multivariable logistic regression model with 16 variables to estimate the probability of mortality for patients admitted to acute care hospitals with severe sepsis or septic shock. The model was built using the development dataset and starting with all possible covariates in the model. Using an iterative procedure, variables were removed from the model, one by one, if the p-values were not significant at 0.05 level until a parsimonious model was reached.
  - Variables removed during the development procedure were added back into the reduced model if the p-values were significant at the 0.05 level and if model calibration (Hosmer-Lemeshow goodness of fit) was improved through their inclusion.
  - The scale of the three continuous variables (patient age, first serum lactate, and the count of the number of comorbidities) remaining in the model was assessed. Using the method of fractional polynomials patient age was included in the model as a linear term, number of comorbidities was transformed by taking the square root, and first serum lactate was entered into the model as a quadratic expression (linear and a squared term).
  - Model calibration was further improved by adding the following interactions to the model: lower respiratory infection (LRI) by MV severity, patient age by the square root of the number of comorbidities, and first serum lactate by the square root of the number of comorbidities.
  - Age, gender, payer, race and ethnicity were initially included in the model. Gender was the only variable included in the model since its odds ratio and corresponding p-value was 1.0003 and 0.992, respectively. All of the other demographic variables had p-values < 0.001 for at least one of the levels of a specific demographic.
- The developer stated that the intent of this risk model is to estimate the *probability* of mortality due to sepsis not *predict* mortality rates due to sepsis. By estimating the probability, the developer continued, the expected number of events (sepsis mortality) for each hospital is calculated. Variables must be clinically and statistically significant to be included in the risk model. The Committee recognized that the developer deliberately excluded hospital characteristics from the variables in the risk model because organizational variables do modify the probability of mortality for patients.
- One of the Committee members stated that they would like to see a statistical analysis, such as a funnel plot, demonstrating outliers (i.e., # of hospitals, hospital size and statistical threshold). A funnel plot is helpful in illustrating real variation among hospitals vs. noise.
- Ultimately, the Committee agreed the measure met the validity criterion.

3. Feasibility: H-1; M-11; L-3; I-1

3215 Adult Inpatient Risk Adjusted Sepsis Mortality
(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented) Rationale: • The developer noted that some data elements are in defined fields in electronic sources, and that some demographic variables can be extracted electronically and used in a standard format. Other variables are collected manually by hospitals though some hospitals have created electronic data capture avenues. • During the workgroup call, the Committee discussed the amount of manual chart abstraction that would be required to collect the data for this measure. The Committee acknowledged the challenges related to the feasibility of this measure but agreed that it was not impossible.
4. Usability and Use: H-2; M-10; L-1; I-3 *(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)* Rationale: • This measure is in use in New York State. New York State requires the Department of Health to collect and report data regarding the performance of hospitals for patients with sepsis including risk adjusted mortality rates for individual hospitals. • The Committee agreed that hospitals in other states could use this measure to track sepsis mortality outcomes.
5. Related and Competing Measures • This measure is related to: o #0500: Severe Sepsis and Septic Shock: Management Bundle • The developer stated the measure specifications are harmonized to the extent possible.
Standing Committee Recommendation for Endorsement: **Y-11**; **N-5**
6. Public and Member Comment • One comment was submitted in support of the Committee's recommendation for endorsement.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Board of Directors Vote: Y-X; N-X
9. Appeals

0500 Severe Sepsis and Septic Shock: Management Bundle

Submission | **Specifications**

Description: This measure focuses on adults 18 years and older with a diagnosis of severe sepsis or septic shock. Consistent with Surviving Sepsis Campaign guidelines, the measure contains several elements, including measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. As reflected in the data elements and their definitions, these elements should be performed in the early management of severe sepsis and septic shock.

Numerator Statement: The number of patients in the denominator who received ALL of the following components (if applicable) for the early management of severe sepsis and septic shock: initial lactate levels, blood cultures, antibiotics, fluid resuscitation, repeat lactate level, vasopressors, and volume status and tissue perfusion reassessment.

- Within 3 hours of presentation of severe sepsis:
    - Measure initial lactate level
    - Draw blood cultures prior to antibiotics
    - Administer broad spectrum or other antibiotics
- Within 6 hours of presentation of severe sepsis:
    - Repeat lactate level (if initial lactate > 2 mmol/L)
- Within 3 hours of presentation of septic shock:
    - Administer 30 ml/kg crystalloid for hypotension or lactate ≥ 4 mmol/L
- Within 6 hours of presentation of septic shock:
    - Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation) to maintain a mean arterial pressure (MAP) ≥ 65 mm Hg
    - Reassess volume status and tissue perfusion in the event of persistent hypotension (MAP <65 mm Hg) after initial fluid administration or initial lactate level ≥ 4 mmol/L
        - The clinician is no longer required to state the method of reassessment used (e.g. physical exam, bedside cardiovascular ultrasound, passive leg raising, CVP, ScV02 assessment). The clinician can attest that volume and perfusion reassessment has occurred, even without reference to the method used. This will meet the measure's volume and perfusion reassessment requirement. A provider may also opt to state their chosen method, but this is not required.

Denominator Statement: Inpatients age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock.

Exclusions: The following patients are excluded from the denominator:

- Severe sepsis is not present
- Patients Transferred in from another acute care facility
- Patients receiving IV antibiotics for more than 24 hours prior to presentation of severe sepsis.
- Patients with a Directive for Comfort Care or Palliative Care within 3 hours of presentation of severe sepsis
- Patients with an Administrative Contraindication to Care within 6 hours of presentation of severe sepsis
- Patients with an Administrative Contraindication to Care within 6 hours of presentation of septic shock
- Patients with a Directive for Comfort Care or Palliative Care within 6 hours of presentation of septic shock
- Patients with septic shock who are discharged within 6 hours of presentation
- Patients with severe sepsis who are discharged within 6 hours of presentation
- Patients with a Length of Stay >120 days
- Patients included in a Clinical Trial

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility

Setting of Care: Hospital

0500 Severe Sepsis and Septic Shock: Management Bundle

Type of Measure: Composite

Data Source: Imaging-Diagnostic, Laboratory, Other, Paper Records, Pharmacy

Measure Steward: Henry Ford Hospital

STEERING COMMITTEE MEETING [03/14/2017]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **H-4**; **M-9**; **L-0**; **I-2**; 1b. Performance Gap: **H-14**; **M-0**; **L-1**; **I-0; 1c. Composite – Quality Construct and Rationale: H-5; M-7; L-1; I-2**

<u>Rationale</u>:

- For the 2012 endorsement-maintenance evaluation, the developer provided the 2008 Surviving Sepsis Campaign guidelines with recommendations for initial resuscitation, measuring lactate, obtaining appropriate blood cultures, antibiotic therapy, fluid therapy, vasopressors and monitoring central venous pressure (CVP) and central venous oxygen saturation ($ScvO_2$).
- In 2012, concerns were raised about the level of evidence supporting invasive monitoring of CVP and ScvO2. The Infectious Disease Steering Committee acknowledged these concerns yet determined that the current evidence at the time was sufficient to warrant endorsement of the full bundle, and the measure was approved as specified. NQF received an appeal and the Consensus Standards Approval Committee (CSAC) upheld the measure's endorsement with the condition that NQF commit to an immediate re-evaluation of the measure upon release of new evidence from several ongoing studies including the Protocolized Care for Early Septic Shock (ProCESS) trial.
- In 2014, the Patient Safety Standing Committee conducted an *ad hoc* review based on a request from the American College of Emergency Physicians (ACEP). The *ad hoc* review focused on the evidence supporting CVP and ScvO2 and the new data from the ProCESS trial. See NQF-Endorsed Measures for Patient Safety (January 30, 2015) for complete summary.
    - The ProCESS trial demonstrated no difference in mortality outcomes when using an invasive approach to monitoring CVP and ScvO2 compared to usual care or protocolized care without invasive monitoring. The Committee noted that the new results from the ProCESS trial suggested that a mandate to measure CVP and ScvO2 with an invasive line might not be necessary in all patients with severe sepsis and septic shock.
    - Experts in support of maintaining these elements in the measure argued additional trials (ARISE and PROMISE) were underway; however, these trials were smaller than the ProCESS trial and not performed in the U.S. In addition, these experts argued that the protocolized care and requirement for CVP and ScvO2 monitoring was particularly helpful in community hospitals, which were not included in the ProCESS trial.
    - After extensive discussions and negotiations the measure developers, ProCESS trial investigators and specialty societies (including the Society of Critical Care Medicine (SCCM) and the American College of Emergency Physicians (ACEP)) reached a compromise for an evidence-based replacement element – optional measurement of CVP and ScvO2, along with reassessment by other means (re-assess volume status and tissue perfusion after initial resuscitation and document findings).
- For the current maintenance of endorsement evaluation, the developer provided the following updated evidence to support the changes to the measure since the last submission:
    - The developer provided a synthesis of the literature for the following updated components, which are based on the Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. ***[Note: Grading of recommendations for the components below are taken from the Sepsis and Septic Shock 2016 Guidelines]***
        - Measure lactate level; Repeat lactate if initial lactate is elevated *[weak recommendation, low quality of evidence[1]]*

[1] Per guideline authors, 'low' grade assigned to quality of evidence (5 RCTs) because 1) all studies were judged to be at high risk of bias due to lack of clarity of the intervention, therefore, we downgraded the quality of evidence by one level for risk of bias; 2) We downgraded the quality of evidence by one level for imprecision, the CI contained small benefit that was lower than the decision threshold; and 3) We assumed a mortality rate for patients with septic shock to be 40%.

- Obtain cultures prior to antibiotics *[Best practice statement]*
- Administer broad spectrum antibiotics *[strong recommendation, moderate quality of evidence]*
- Administer 30 ml/kg crystalloid for hypotension or lactate ≥ 4 mmol/L *[strong recommendation, low quality of evidence]*
- Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation to maintain a mean arterial pressure ≥ 65 mmHg) *[strong recommendation, moderate quality of evidence]*
- Reassess volume status and tissue perfusion *[Best practice statement].* The developer provided a synthesis of the literature for some common practices used when reassessing volume status and tissue perfusion in the event of persistent hypotension (MAP <65 mm Hg) after initial fluid administration or initial lactate level ≥ 4 mmol/L. ***[Note: Clinician is no longer required to document the method used; attestation is sufficient]***

- One of the Committee members questioned the use of lactic levels to diagnose sepsis because other conditions can elevate lactate levels. The developer responded that measuring lactate levels alone does not diagnose sepsis. However, in the presence of a suspected infection, elevated lactate levels are useful in determining illness severity. The Committee pointed out the varying level of evidence for the different components but agreed that overall, the updated evidence is consistent with the 2016 Guidelines for Management of Sepsis and Septic Shock.
- The developer provided the following composite performance rates from CMS' Hospital Inpatient Quality Reporting (IQR) program from October 2015 to June 2016:

	Oct-Dec 2015	Jan-Mar 2016	Apr-Jun 2016
# of hospitals	3,134	3,182	3,193
# of eligible cases	96,516	104,166	101,599
Overall performance rate	34.4	39.5	44.0
10th percentile	5.0	7.7	12.5
25th percentile	17.9	21.6	25.8
Median	31.0	36.1	41.7
75th percentile	45.8	51.3	57.1
90th percentile	60.0	66.7	71.4
Min, Max	0.0, 100.0	0.0, 100.0	0.0, 100.0
Average	32.6	37.1	41.9
Standard Deviation	21.1	21.9	22.9

- The developer also provided the following component rates categorized by 3 and 6 hour elements:

Population Description	Cases	Bundle %
Initial Population Number of Sepsis Cases	**325,809**	**--**
Total Number of Excluded Sepsis Cases	*166,520*	--
Total Number of Eligible Sepsis Cases	**159,289**	**--**
Total Number of Passed Sepsis Cases	*64,051*	--
Total Number of Failed Sepsis Cases	*95,238*	--
Severe Sepsis 3 Hour Bundle Eligible Cases	**167,114**	**--**
Severe Sepsis 3 Hour Bundle Passes	*110,078*	*65.9%*
Severe Sepsis 3 Hour Bundle Failures	*54,618*	*32.7%*
Initial Lactate Level Failures	26,503	48.5%
Broad Spectrum or Other Antibiotic Administration Failures	20,951	38.4%

Blood Culture Collection Failures	18,772	34.4%
Severe Sepsis 6 Hour Bundle Eligible Cases	**90,385**	--
Severe Sepsis 6 Hour Bundle Passes	*53,475*	*59.2%*
Severe Sepsis 6 Hour Bundle Failures	*36,910*	*40.8%*
Repeat Lactate Level Failures	36,910	
Septic Shock 3 Hour Bundle Eligible Cases	**40,989**	--
Septic Shock 3 Hour Bundle Passes	*22,359*	*54.5%*
Septic Shock 3 Hour Bundle Failures	*18,630*	*45.5%*
Crystalloid Fluid Administration Failures	18,630	
Vasopressor Shock 6 Hour Bundle Eligible Cases	**8,177**	--
Vasopressor Shock 6 Hour Bundle Passes	*6,157*	*75.3%*
Vasopressor Shock 6 Hour Bundle Failures	*2,020*	*24.7%*
Vasopressor Administration Failures	2,020	
Focus Exam Shock 6 Hour Bundle Eligible Cases	**14,630**	--
Focus Exam Shock 6 Hour Bundle Passes	*3,801*	*26.0%*
Focus Exam Shock 6 Hour Bundle Failures	*9,935*	*67.9%*
Hemodynamic Choices Shock 6 Hour Bundle Eligible Cases	**10,829**	--
Hemodynamic Choices Shock 6 Hour Bundle Passes	*894*	*8.3%*
Hemodynamic Choices Shock 6 Hour Bundle Failures	*9,935*	*91.7%*

- The developer noted that the 'repeat volume' and 'perfusion assessment' data is broken down into 'focused exam' and 'hemodynamic choice' – data elements that are no longer required. No performance data is yet available on the new attestation strategy.
- The developer also provided the following composite performance rates by ethnicity, gender and Medicare/non-Medicare:

	Oct-Dec 2015	Jan-Mar 2016	Apr-Jun 2016
Hispanic	34.53	39.8	44.23
Non-Hispanic	32.64	36.35	40.82
Females	33.82	39.22	43.28
Males	34.93	39.86	44.63
Medicare	34.9	40.07	44.57
Non-Medicare	33.32	38.47	42.76
Black	30.64	35.93	40.29
White	34.95	40.08	44.58
Other	34.95	40.01	43.82

- The performance rates for different age categories were similar (~34.0).
- The Committee agreed that the developer presented abundant data demonstrating a performance gap and opportunity for improvement in severe sepsis and septic shock care.
- Another Committee member commented that in addition to a performance gap in care, there is still a lack of implementation of this measure in the clinical setting.
- This all-or-none composite measure requires patients with severe sepsis or septic shock to meet all of the eligible components in the composite: lactate collection, delivery of broad-spectrum antibiotics, obtaining blood cultures, delivering resuscitation fluids, applying vasopressors as needed, reassessing volume and perfusion status and repeating lactate values. All components are equally weighted. The developer noted that the composite ensures a strategy aimed at reducing mortality. The components could not

stand alone unless certain preceding conditions have been met. In addition, the components are aggregated in three and 6-hour elements for severe sepsis and septic shock.

- A Committee member suggested that the data elements should be weighted differently based on the level of evidence for each. The Committee member also questioned the two different time periods. The developer responded that severe sepsis and septic shock are two different diagnoses that qualify for this measure. Therefore, the measure is constructed so that there are dependencies, both in time and condition, based on the diagnosis of severe sepsis or septic shock. The severity of the patient's condition determines whether they qualify to receive all of the components in the composite measure and when. The Committee agreed, that overall, the quality construct and rational for the composite was clearly stated and logical**.**

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-7**; **M-7**; **L-0**; **I-1** 2b. Validity: **H-0**; **M-11**; **L-1**; **I-3**

<u>Rationale</u>:

- For the 2012 endorsement evaluation, the developer conducted a signal-to-noise analysis of the individual bundle elements and overall bundle reliability and composite measure reliability by site for 165 hospitals and 15,022 patients from January 2005 – March 2008.
- Effective 2013, NQF determined that reliability of the individual component measures was not sufficient; reliability must be demonstrated for the composite measure score.
- For the current evaluation, the developer provided updated reliability testing at the composite score level using a random sample of SEP-1 chart-abstracted data submitted to CMS as part of the Hospital Inpatient Quality Reporting (IQR) program. The sample included 302,281 cases in the denominator (after exclusions) and 119,048 cases in the numerator from 3,134 to 3,193 hospitals (depending on the quarter) from October 2015 to June 2016.
- The developer used a beta-binomial model to assess the signal-to-noise ratio. A reliability of 0.0 implies that all the variability in a measure is attributable to measurement error. A reliability of 1.00 implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one provider from another. This is an appropriate test for measure score reliability. A reliability of 0.70 is generally considered a minimum threshold for reliability.
    - The median reliability score was calculated including all facilities and facilities with a minimum of 10 eligible cases (more than 86% of reporting facilities).
    - The developer provided the overall reliability score for the composite measure for each quarter:
        - October 2015 – December 2015: 0.92 (CI 0.41 – 1.00)
        - January 2016 – March 2016: 0.93 (CI 0.47 – 1.00)
        - April 2016 – June 2016: 0.93 (CI 0.42 – 1.00)
- In the pre-evaluation comments, the Committee expressed some concerns that the self-reported clinician attestation for the data element 'volume and perfusion reassessment' may lead to more subjectivity but lauded the developers in their efforts to reduce documentation and chart abstraction burden.
- During the in-person meeting, the Committee did not voice any additional concerns related to reliability and agreed the reliability testing results were sufficient.
- For the 2012 endorsement evaluation, the developer assessed the validity of the measure score by testing the hypothesis that those with higher scores on the composite performance measure should have a lower score on a risk-adjusted mortality measure. The developer reported that hospital mortality was reduced by 10% for patients that were compliant with the composite measure.
- For the current evaluation, the developer provided updated validity testing at the composite score level. The developer performed a Chi-Square Test of Association and Equal Proportions between two categorical variables: Measure Outcome (Failed or Passed) and Mortality Result (Died or Survived). The Chi-Square Test of Association and Equal Proportions demonstrated a p-value of <.0001, a risk ratio of 1.3856, a lower 95% confidence limit of 1.3616 and an upper 95% confidence limit of 1.4101. A risk ratio higher than 1.0 with a significant p-value, would indicate that there is a higher risk of dying when a case fails the measure compared to when a case passes the measure.

  - The risk ratio can be expressed as an actual ratio and it can be said with 95% confidence, cases that fail the measure have 1.36 to 1.41 times the risk of dying compared to cases that pass the measure; or
  - The risk ratio can also be used as a percentage and be said that with 95% confidence, cases that fail the measure have a 36% to 41% increase in risk of dying compared to cases that pass the measure.
- The developer provided sepsis rate comparisons analyses, which demonstrated a negative association between pass rates and mortality rates from October 2015 to June 2016.
- The developer also included an analysis of pass rates and mortality rates by percentiles and a two-proportions z-test. The z-test determines if there is a statistically significant difference in mortality rates between percentiles. These methods are appropriate for empirically assessing the validity of the composite measure score. The results of the sepsis mortality analysis demonstrated that 30.4% of the total number of 'Failed Sepsis Cases' died (at discharge and up to 30 days after discharge) compared to 21.9% of the total number of 'Passed Sepsis Cases'. The two-proportion z-test demonstrated that four of the percentile comparisons have a statistically significant difference between mortality rates at a significance level of 0.05. Three additional percentile comparisons are fairly close to a statistically significant difference between mortality rates at a significance level of 0.10.
- The developer also provided the mortality rate for patients who received all applicable elements of care for the composite measure (passed sepsis cases) and those who did not (failed sepsis cases) for each quarter – the mortality rate for those who received all applicable elements was on average 8.5% lower compared to patients who did not receive all applicable elements of care.

	Severe Sepsis and Septic Shock Mortality Rate		
Description	**2015 Q4**	**2016 Q1**	**2016 Q2**
Did not Meet Guidelines for SEP-1	29.6%	31.8%	29.7%
Met Guidelines for SEP-1	21.3%	23.0%	21.4%
Absolute Reduction Rate	8.3%	8.8%	8.3%
Relative Reduction Rate	28.04%	27.7%	27.9%
	Potential Preventable Deaths		
	2,783	2,864	2,411

- The Committee discussed the results of the patient-level data element validity testing conducted by CMS. Several Committee members were concerned because out of the 55 data elements tested for validity, 15 data elements (27.27%) had a percent agreement higher than 90%. The remaining 40 data elements (72.73%) had a percent agreement lower than 90%. The developer stated that there have been numerous education and outreach efforts and updates to the measure with the intent of clarifying guidance and decreasing abstractor complexities in an effort to improve successive validation testing. NQF staff noted that this method is not appropriate for composite measures. NQF composite performance measure evaluation guidance (2013) states that validity testing is directed toward the inferences that can be made about accountable entities on the basis of their performance measure scores. For the purposes of endorsing composite performance measures, validity testing of the constructed composite performance measure score is more important than validity testing of the component measures. Even if the individual component measures are valid, the aggregation and weighting rules for constructing the composite could result in a score that is not a true reflection of quality (p. 12-13).
- The developer stated that the number of exclusions was not significant enough to unfairly distort measure performance results and potentially negatively affect the reliability of the measure because the vast majority of the exclusions were cases where severe sepsis was not present (72.34%) and should not be analyzed.
- The developer provided the table included in performance gap to demonstrate the contribution of each component to the composite score.
- The Committee agreed that the measure met the validity criterion.

0500 Severe Sepsis and Septic Shock: Management Bundle

3. Feasibility: H-1; M-9; L-5; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- Data elements are abstracted from a record by someone other than the person obtaining the original information (e.g. chart abstraction); some data elements are in defined fields in electronic sources.
- The developer noted that the measure is complex and requires data abstractors to "comb through documentation and interpret" clinician documentation; however, the most recent updates to the measure should lessen documentation and abstractor burden. The developer also stated that the measure has gone through three updates to lessen abstractor burden and address issues related to data availability, missing data, and frequency of data collection.
- The developer also stated that preliminary efforts to convert this measure to an electronic measure within the HQMF/QDM framework were not feasible. Currently, there are no plans to respecify this measure into an eMeasure.
- There are no fees or licenses required to use this measure.
- In the pre-evaluation comments and during the in-person meeting, the Committee noted that the feasibility of this measure was a significant concern. In the pre-evaluation comments one of the Committee members acknowledged that although clinicians routinely document the required data elements for this measure, it is complex and likely results in significant time and costs associated with data collection and reporting. During the in-person meeting, the developer clarified that it is the clinician's responsibility to document appropriate care, therefore, decreasing the burden on the abstractors.
- The developer and CMS representatives reiterated that they regularly receive feedback from hospitals and their abstractors and are currently monitoring the most recent changes to the specifications and implementation guide.
- The Committee concluded that the feasibility of this measure is challenging but it meets the criteria.

4. Usability and Use: H-2; M-6; L-6; I-0 Consensus was not reached on the Usability and Use criteria

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The measure is currently used in CMS' Hospital Inpatient Quality Reporting (IQR) Program for acute care hospitals nation-wide. Across the three quarters of available data, between 3,134 and 3,193 providers submitted data, which represents more than 95% of eligible providers nationwide.
- The measure is not currently publicly reported, but will be added to the Hospital Compare website at a date to be determined. Due to the complexity of the measure specifications, CMS desires to review and analyze the data prior to making it publicly available. There were also several updates to the specifications based on stakeholder feedback, and CMS wants to assure stability of the specifications before public reporting.
- The Committee discussed the unintended consequences associated with diagnosing and treating sepsis. Because sepsis is a combination of symptoms rather than a disease, the Committee noted, patients often receive antibiotics prior to a conclusive diagnosis of sepsis – this is similar to the unintended consequences that occurred with the overuse of antibiotics in the emergency room in an effort to meet the now retired pneumonia measure.
- At the end of the discussion, the Committee did not reach consensus on the usability and use of the measure during the in-person meeting.

5. Related and Competing Measures

- This measure is related to:
    - 3215: Adult Sepsis Mortality Outcome Measure (New York State Department of Health)

Standing Committee Recommendation for Endorsement: **Y-10**; **N-4**

6. Public and Member Comment

Nine comments were submitted covering the following themes: antibiotic administration, level of evidence, and scientific acceptability.

- In regard to antibiotic administration, commenters noted the three hour time window for antibiotic administration may lead to antibiotic overuse; other comments questioned whether antibiotic administration rather than early goal directed therapy had a larger effect on patient survival rates; and another comment suggested that the developer reduce the antibiotic administration time window from three hours to two hours.
    - Developer Response to Comment ID 6678: We understand that measures can have unintended consequences for patients. We remain convinced that in severe sepsis (please refer to clinical criteria used for the measure) and septic shock the risk of mortality is so high that is critical to provide early and broad antibiotic therapy. The failure to provide a proper antibiotic for a patient with severe sepsis and septic shock carries a much higher risk than the potential harm of providing a single dose of a broad spectrum antibiotic to a patient who turns out not to have severe sepsis or septic shock. This in by no means precludes the clinician's responsibility to judiciously use IV antibiotics in a way which upholds the standards of antibiotic stewardship. The 2016 Surviving Sepsis Guidelines, endorsed by the Infectious Disease Society of America speak to this question: "The rapidity of [antimicrobial] administration is central to the beneficial effect of appropriate antimicrobials. In the presence of severe sepsis or septic shock, each hour delay in administration of appropriate antimicrobials is associated with a measurable increase in mortality." When mortality already approaches 18-40% in shock states, it is unacceptable to suspend antibiotic administration pending further studies. However, in the event an antibiotic is given inappropriately in non-sepsis states, the guidelines also recommend, "Given the potential harm associated with unnecessarily prolonged antimicrobial therapy, daily assessment for de-escalation of antimicrobial therapy is recommended in patients with severe sepsis and septic shock."
    - Developer Response to Comment ID 6680: The PRISM investigators have reported results in the New England Journal of Medicine (NEJM) entirely consistent with the prior Process, Promise and Arise trials, also published in NEJM. Little information is imparted by PRISM that was not already known from these previous trials. In fact, PRISM derived all data from the prior trials. We emphasize that SEP-1 is consistent with the conclusions of these trials and does not require an invasive method of patient reassessment. In regards to Dr. Kalil's publication in Critical Care Medicine, "Early Goal-Directed Therapy for Sepsis: A Novel Solution for Discordant Survival Outcomes in Clinical Trials," the major conclusion was that: "[S]urvival discordance was not associated with differences in early goal-directed therapy bundle compliance or hemodynamic goal achievement. Our results suggest that it was associated with faster and more appropriate antibiotic co-intervention in the early goal-directed therapy arm compared with controls in the observational studies but not in the randomized trials." While we may dispute the methods and analysis used to reach this conclusion, we again underscore that SEP-1 does not mandate an invasive reassessment but does require early IV antibiotic administration. Thus SEP-1 is consistent with the Kalil publication in its approach.
    - Developer Response to Comment ID 6806: The developers will take the Armstrong Institute's helpful suggestion under advisement and model data to understand how a 2 hours standard would affect the performance characteristics of the measure. We agree that earlier administration of antibiotics is the preferred approach.
- Comments also noted the varying level of evidence for the different components in the measure composite: repeat lactate, fluid reassessment, and physical exam. The comment suggested that these components are not equivalent to antibiotic or IV fluid administration and should not be weighted equally in the composite construct. Another comment recommended a simplified three hour bundle without the repeat lactate and physical exam component.
    - Developer Response to Comment ID 6708: We will take under consideration the suggestion to weight elements in different ways than presently weighted. It is important to note however that as a matter of process for vetting measures, they must be advanced on the basis of accumulated data and evidence. SEP-1 continues to show a high association with reduction in mortality with the individual specified elements as documented in the submission which justified the weighting. As the submission shows, there is a large separation in mortality between those who comply with the elements in total and those who fail any one or more than one of the elements. To understand and model a proposal such as Dr. Doerfler's will require analysis of the measure as a

component measure, which necessarily means analyzing the data in a fashion for which it was not designed. We will discuss with CMS Dr. Doerfler's hypothesis regarding preferential weighting of certain data elements.

- Developer Response to Comment ID 6810: The developers appreciate the opportunity to respond to the High Value Healthcare Collaborative (HVHCC) which has advanced excellent work in quality improvement and care of patients with severe sepsis and septic shock. We note that the HVHCC has indicated that SEP-1 endorses an "all-or-none payment approach." SEP-1 performance does not impact payment to hospitals or providers. As regards concerns that SEP-1's composite construction does not differentiate between importance of antibiotic administration and a physical exam element (cited as skin color), the developers would like to point out that, in the current specification manual and in this NQF submission, SEP-1 does not require documentation of particular physical exam elements. A provider may now indicate simply (within the allotted timeframe) that they have "performed a physical exam" without regard to a means or method, physical exam or otherwise, and pass this data element. In this regard, the developers would suggest that regular reassessment of a patient with septic shock as regards to perfusion status is as important as antibiotic administration and supports the composite construction as advanced in this submission.

As regards the representation that "once a mature care model is in place, compliance with a 3-hr-bundle, had no impact on in-hospital, 30-day, 90-day or 1-year post discharge mortality between those receiving the full bundle vs not. This is consistent with the conclusions from the ProCESS, ARISE, and ProMISE trials," this claim is not an accurate representation of the cited trials. The 3 hour elements of care were required of every patient in the cited trials. All patients received the three hour elements (initial lactate, blood culture collection, and broad spectrum antibiotic administration) prior to randomization. Aside from this inaccuracy, it is unclear what the characteristics of a "mature care model" may be in the HVHCC's remarks, however generalizing that the 3000+ hospitals in the United States subject to SEP-1 have such a model is unsupported by any evidence to properly analyze that claim. Moreover, since HVHCC does endorse a "simplified 3-hour-bundle" it would seem to remain an important element of care. The developers do not believe that clinicians and hospitals may not define "innovative approaches to early sepsis detection" under SEP-1. One method to ascertain time zero that overrides all other methods is a provider's documentation of the time. In that regard, the HVHCC may use whatever method they prefer to set a time zero as long as their clinicians concur with the HVHCC's approach. The developers appreciate the HVHCC's suggestion to proceed with the 3 hour elements in SEP-1 and we assure them that these elements remain in this submission. As regards the representation that there is "no evidence" supporting repeat lactate assessment or a physical exam, the developers repeat that clinicians must only document reassessment of perfusion or volume status by any means of their choosing. This practice, along with repeat lactate assessment do have a supporting evidence base in the 2016 Surviving Sepsis Campaign guidelines. Reassessment is a best practice statement under the GRADE evaluation framework and repeat lactate assessment has a low quality of evidence with a weak recommendation under the same criteria. We will take under advisement that these elements should be examined further in future iterations of the specifications and would welcome the opportunity to work with the HVHCC to model these approaches. We note that the measure does not apply to critical access facilities and is not active at this time in any pay for performance programs.

- Two other comments questioned the percentage agreement rates among the patient level data elements. The commenters also disagreed with the guidance given to the standing committee on evaluation composite measures.
  - Developer Response to Comment ID 6803-6804: The Federation of American Hospitals has raised many questions that were previously discussed in detail in committee. We appreciate the opportunity to summarize these issues. While burden of data collection may be greater than for other measures in healthcare, this is more than counterbalanced by severe sepsis and septic shock's' burden on the healthcare system as the number one cause of inpatient deaths in the United States and highest cost condition for hospital admissions. Evidence that the SEP-1 measure drives quality improvement was provided at NQF. The initial three quarters of data analysis show that hospitals improved their performance from quarter to quarter. In addition, the analysis revealed a statistically significant finding that there is an approximately 8.5%

associated reduction in mortality in those who comply with the measure versus those who fail the measure. As measure developers we have no data to comment on the quality of responses provided by Quality Net, but we will share the feedback with the Quality Net team. We cannot address the Federation's representations regarding the motives of other agencies to utilize the measure, but we have provided substantial data that SEP-1 meets the standards set out in NQF's measure evaluation framework. As regards validity, the measure met standards for assessing validity at the performance score level, which is the proper level of evaluation for a composite measure. Additionally, it is precisely for the Federations' argument of the limited sample size (303 cases) that the data element level validity cannot serve as a valid critique of SEP-1. In addition, the measure met all reliability criteria with statistically appropriate analysis using a signal to noise methodology. While the Federation states that the element level validity testing is more important than the performance measure score testing, under the NQF measure evaluation framework, that choice is an improper standard to evaluate a composite measure. We note the Federation misinterprets the Technical Expert Panel's remarks that "the individual components may not be sufficiently reliable independently, but could contribute to the reliability of the composite performance measure." First, this quotation refers to reliability whereas the Federation was addressing validity. Secondly, the principle that the individual elements could contribute to the overall validity at the level of the performance score is precisely the point of the Technical Expert Panel's comment. This rationale is why element validity is not the criterion for composite measures. Finally, the Federation has advanced no evidence to evaluate the claim that the measure does not meet the validity criteria set out by NQF at the performance measure score level. The Federation has inadvertently misstated the evaluation criteria. Specifically, subcriterion 2d states that "[f]or composite performance measures, empirical analyses support the composite construction approach…" Nothing in the Federation's remarks indicates that the composite construction approach is under question. As regards "missing data," the measure evaluation frame work considers this under subcriterion 2b7, which requires that the measure "analyses and identifies the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders and how the specified handling of missing data minimizes biases." This requirement is different than the scenario described by the Federation which equates the variation in element level validity testing with missing data. This equivalency is incorrect insofar as the analysis of the complete data shows that some degree of variation actually exists – if the data were "missing" a showing of variation could not be made. In this regard, the developers stand by their submission that all data is present and cannot be missing as part of the reporting requirements met by 99.9% of over 3200 acute care hospitals in three consecutive quarters of analyzed data. To the extent that SEP-1 data elements may be contained in an electronic health record, the developers will take the Federation's excellent suggestion under advisement to consider, if feasible, the measure as an e-measure. The Federation implies that hospitals do not understand why they fail the measure given its composite nature. The developers believe, however, that by analyzing the point of failure in the measure framework, providers know exactly where weak spots are in their clinical care processes. For example, failure of the antibiotic element indicates a process in antibiotic administration that needs attention. In addition, the software provided by major vendors to hospitals to report the measure specifically categorizes the level of the fallout for facilities. The developers strenuously object to the characterization that the measure has limited value in improving patient care: in over 600,000 patients captured by the measure there has been an approximately 8.5% reduction in mortality in those compliant with the measure versus those who were not. For the population of submitted cases, this represents a potential lives saved calculation of over 7,500 patients in the first three quarters of data for the measure.

- Developer Response to Comment ID 6811-6813: The developer's appreciate the opportunity to respond to the remarks of the American Medical Association (AMA) and clarify the operation of SEP-1. As regards to Dr. Pronovost's publication in the American Journal of Medical Quality regarding possible unintended consequences of quality measures, we note that Dr. Pronovost is the Director of the Armstrong Institute for Patient Safety and Quality which has recommended that SEP-1 be re-endorsed with the exception of tightening the antibiotic administration

requirement from 3 hours to 2 hours. Please see the submitted comment of Dr. Matt Austin, PHD, on behalf of the Armstrong Institute for Patient Safety and Quality at Johns Hopkins University, which was also received during the post-evaluation comment period. With respect to the specific concern that fluid administration as specified in SEP-1 may be harmful to patients with left ventricular systolic dysfunction, the AMA cites an opinion article authored by an emergency medicine resident. We note that this opinion is not representative of any clinical trial, observational or randomized, controlled or not. The opinion cites another article by Boyd 2011 which indicates that a positive fluid balance and elevated CVP are associated with increased mortality. The developers note that SEP-1 does not advocate for a "positive fluid balance" which refers to volume status over several days of care. SEP-1 is limited to initial resuscitation and the first 6 hours of care after presentation. In addition, SEP-1 does not advocate for CVP measurement, and certainly not an "elevated CVP." Another citation in the resident's article is Pudilo 2012 which reports frequency of myocardial dysfunction in severe sepsis and septic shock. The article by Pudilo actually points to a reason for proper volume resuscitation of patients: the well-known presence of severe sepsis induced myocardial dysfunction. The salient finding is not that the patients have intrinsic heart disease, but rather that sepsis has caused impaired myocardial function. In addition, Pudilo does not conclude that a 30 ml/kg initial fluid bolus as an initial resuscitation strategy in septic shock is detrimental to patients. This Pudilo reference does not support any of AMA's criticism of the SEP-1 measure. On the broader concern about the potential risks of fluid resuscitation, we note that there is no published evidence from any randomized controlled trial which indicates that patients with severe sepsis and known congestive heart failure or renal failure who receive a fluid bolus for initial resuscitation do worse than other sepsis patients in terms of mortality. In fact, even the examination of the large trials on septic shock do not support this contention (EGDT 2001, Process 2014, Promise 2014, Arise 2015). In fact the only published evidence on the topic concludes that for patients with intermediate lactate vales of 2-4 mmol/L who receive the full fluid bolus of 30 ml/kg with congestive heart failure and renal failure have lower mortality than their counterparts without these co-morbidities. (See: Lui V et al. Fluid Volume, Lactate Values, and Mortality in Sepsis Patients with Intermediate Lactate Values. Ann Am Thorac Soc Vol 10, No 5, pp 466-473, Oct 2013). The thrust of the AMA's comments on this topic regarding LVSD is that physician judgment should be preserved. The developers agree with the AMA that physician judgment is paramount and agree that providers should exercise their best judgment informed by the evidence when caring for sepsis patients. SEP-1 is not a prescriptive recipe for all patients with severe sepsis and septic shock, but rather a measurement strategy for processes of sepsis care. The developers fully expect that practitioners will do what is best in their understanding for each patient, which may result in deviation from SEP-1. Ultimately when sufficient data is amassed, best compliance, which takes into account those necessary deviations, will be known. In that regard, there is no expectation or goal of a 100% compliance with SEP-1. However, we emphasize that all current evidence suggests better mortality with higher compliance. As regards to the concern that a precisely specified measure should account for how unplanned drug shortages could impact an individual hospital's performance and the concern that mortality varied with a shortage of nor-epinephrine, we note that SEP-does not require the use of any one particular vasopressor. We also note that SEP-1 is a process measure, not an outcome measure such as mortality. Finally, we note that it is likely that all hospitals would be affected by any shortage in a short period of time. Turning to Dr. Kalil's publication in Critical Care Medicine, "Early Goal-Directed Therapy for Sepsis: A Novel Solution for Discordant Survival Outcomes in Clinical Trials," the major conclusion was that "[s]urvival discordance was not associated with differences in early goal-directed therapy bundle compliance or hemodynamic goal achievement. Our results suggest that it was associated with faster and more appropriate antibiotic co-intervention in the early goal-directed therapy arm compared with controls in the observational studies but not in the randomized trials." While we may dispute the methods and analysis used to reach this conclusion, we again underscore that SEP-1 does not mandate Early Goal Directed Therapy and SEP-1 does require early antibiotic administration. Thus SEP-1 is consistent with the Kalil publication in its approach. The PRISM investigators have reported results in the New England Journal of Medicine (NEJM) entirely consistent with the Process, Promise and Arise

trials, also published in NEJM. Little information is imparted by PRISM that was not clearly know from the three other trials, and in fact PRISM derived all data from the prior trials. In any case, SEP-1 as specified is consistent with the conclusions of these trials. AMA has stated concerns with SEP-1's measure performance characteristics. For its validity, the measure met standards for assessing validity at the performance score level, which is the proper level of evaluation for a composite measure. In addition, the measure met reliability criteria with statistically appropriate analysis using a signal to noise methodology. While AMA represents that element level validity testing is more important than the performance measure score testing, this is an improper standard to evaluate a composite measure under the NQF measure evaluation framework. Of substantial importance, we note that AMA misinterprets the Technical Expert Panel's remarks that "the individual components may not be sufficiently reliable independently, but could contribute to the reliability of the composite performance measure." First, the quoted principle that the individual elements could contribute to the overall validity at the level of the performance score is precisely the point of the Technical Expert Panel's recommendation not to focus on element level testing for a composite metric. Second, the AMA has advanced no evidence to evaluate the claim that the measure does not meet the validity criteria set out by NQF at the performance measure score level. In the evaluation of a metric, it would be improper to move the goal post mid-evaluation. Regarding their other comments on validity, the AMA incorrectly cites subcriterion 2d. Subcriterion 2d states that "[f]or composite performance measures, empirical analyses support the composite construction approach…" Nothing in the AMA's remarks indicates that the composite construction approach is under question. In regard to "missing data," the measure evaluation frame work actually considers this under subcriterion 2b7, which requires that the measure "analyzes and identifies the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders and how the specified handling of missing data minimizes biases." This requirement is different than the scenario described by AMA which equates the variation in element level validity testing with missing data. This equivalency is incorrect insofar as the analysis of the complete data shows that some degree of variation actually exists – if the data were "missing" a showing of variation could not be made. In this regard, the developers stand by their submission that all data is present and cannot be missing as part of the reporting requirements met by 99.9% of 3225 acute care hospitals in three consecutive quarters of analyzed data. In summary, the developers appreciate the AMA's comments and suggestions. We agree with the AMA that this disease which places an unacceptable burden in terms of deaths and expenditures on the healthcare system deserves a rigorous and robust measure. For this reason we are proud that SEP-1 (which has over 600,000 reported cases since inception) has an associated 8.5% reduction in mortality for those cases that were compliant with the measure versus those which were not. We look forward to tracking the ongoing impact of the SEP-1 measure on sepsis quality improvement and will share our findings with all concerned stakeholders as they become available.

- During the post comment call, the Committee re-iterated the fact that lactate clearance is an important step in evaluating the patient's condition. The Committee also noted that the intent of reassessment is to determine how well the patient is doing and is a best practice in that it drives collaboration among the care team. In regard to antibiotic administration, the Committee recommended that the developer update the rationale in the measure submission form to include that the intent of the measure is to encourage early administration of antibiotics, preferably within one hour diagnosis of sepsis. The developer committed to testing antibiotic administration within one hour of diagnosis of sepsis. NQF also provided additional guidance to the Committee on the scientific acceptability requirements for a composite measure. The Committee then agreed that the measure still met reliability and validity requirements.
-

7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

8. Board of Directors Vote: Y-X; N-X

9. Appeals

Infectious Disease, 2016-2017

Consensus Standards Approval Committee Review and Recommendations

July 11-12, 2017

Woody Eisenberg, Co-Chair
Adam Thompson, Co-Chair

Melissa Mariñelarena, Senior Director, NQF
Christy Skipper, Project Manager, NQF
Mauricio Menendez, Project Analyst, NQF

NATIONAL QUALITY FORUM

Infectious Disease

- For this project, the Committee evaluated nine measures against NQF's standard evaluation criteria – four new measures and five measures undergoing maintenance review.

- Evaluated measures in the following areas:
  - HIV/AIDS
  - Sepsis and Septic Shock

Infectious Disease

- Recommended measures:
  - 2082: HIV Viral Load Suppression
  - 3210: HIV Viral Load Suppression – Legacy eMeasure
  - 2079: HIV Medical Visit Frequency
  - 3209: HIV Medical Visit Frequency – Legacy eMeasure
  - 2080: Gap in HIV Medical Visit
  - 2083: Prescription of HIV Antiretroviral Therapy
  - 3211: Prescription of HIV Antiretroviral Therapy – Legacy eMeasure
  - 0500: Severe Sepsis and Septic Shock: Management Bundle
  - 3215: Adult Inpatient Risk Adjusted Sepsis Mortality

Infectious Disease

- Recommended measures:
  - **HIV/AIDS – 7 measures**
    - Types of measures: 5 process; 0 composite; 2 outcome
  - **Sepsis and Septic Shock – 2 measures**
    - Types of measures: 0 process; 1 composite; 1 outcome

Infectious Disease

	Maintenance Measures	New Measures	TOTAL Measures
Submitted	5	4	9
Measures Recommended	5	4	9

Overarching Issues:

- Unintended Consequences
  - The Committee expressed concern with the potential unintended consequences associated with diagnosing and suspected sepsis
    - » Combination of symptoms associated with sepsis may be associated with other diagnoses (e.g. pneumonia or myocardial infarction)
    - » Patients may receive antibiotics before a conclusive diagnosis of sepsis
    - » Patients may experience hypotension due to an acute MI rather than sepsis, yet receive IV fluids in order to meet the data elements of the sepsis bundle
- Balancing Measures
  - The Committee expressed the need for balancing measures to avoid unintended consequences:
    - » Overuse of broad spectrum antibiotics
    - » Patients with heart failure and/or an MI who were overloaded with IV fluids
    - » Incidence of C. difficile secondary to overuse of antibiotics

Comments Received

Comments Received: #0500 *Severe Sepsis and Septic Shock: Management Bundle* (Henry Ford Hospital)

- Nine comments received from five NQF member organizations and members of the public

Theme: Antibiotic Administration

- **Summary of Comments**: Comments note that the 3-hour time window for antibiotic administration might lead to unintended consequences like antibiotic overuse. Two commenters questioned whether antibiotic administration rather than early goal directed therapy (EGDT) had a larger effect on patient survival rates as suggested by Kalil et al. (2017). One commenter suggests reducing the antibiotic administration time from three hours to two.
- **Developer Response**: The developers understood the concerns related to unintended consequences of antibiotic overuse, however, "in severe sepsis…and septic shock the risk of mortality is so high that it is critical to provide early broad antibiotic therapy". The developer also noted that the measure "does not mandate an invasive reassessment but does require early IV antibiotic administration" and therefore is consistent with Kalil et al. (2017). Finally, the developers agreed that the earlier administration of antibiotics is the preferred approach and noted they would consider how a 2 hour standard would affect the measure.

Comments Received: #0500 *Severe Sepsis and Septic Shock: Management Bundle* (Henry Ford Hospital)

- **Theme:** Level of Evidence
  - **Summary of Comments:** Comments note the varying level of evidence for the different components in the measure composite. The comments suggest that the level of evidence supporting repeat lactate, fluid reassessment, and/or physical exam is not equivalent to antibiotic or IV fluid administration. The comments state that the components should not be weighted equally in the construct of the composite measure. One commenter also recommends a simplified 3-hour bundle without the repeat lactate and physical exam component.
  - **Developer Response**: The developer will consider preferential weighting of the elements (repeat lactate, fluid reassessment, physical exam). The developer also clarified that although the measure specifications do not require documentation of "particular physical exam elements" that it is just as important as antibiotic administration and therefore supports the measure composite. The developer agreed to further examine reassessment of perfusion or volume status and repeat lactate in future iterations of the measure specifications.

Comments Received: #0500 *Severe Sepsis and Septic Shock: Management Bundle* (Henry Ford Hospital)

- **Theme:** Scientific Acceptability
  - **Summary of Comments**: Comments questioned the patient level data element percentage agreement rates. The commenters also disagreed with the guidance given to the Standing Committee on evaluating composite measures. NQF criteria states that for composite measures, validity and reliability should be empirically demonstrated at the measure score level. The updated reliability and validity composite score level testing provided by the developer meet NQF criteria for composite measures as indicated during the in-person meeting and in the draft report.
  - **Developer Response**: The developer stated that the measure met NQF criteria for scientific acceptability as testing was completed at the measure score level.
  - **NQF Response**: According to NQF composite performance measure evaluation guidance, validity testing of the constructed composite performance measure score is more important than validity testing of the component measures. Even if individual component measures are valid, aggregation and weighting rules for constructing the composite could result in a score that is not a true reflection of quality. The updated reliability and validity composite score level testing meet NQF criteria for composite measures.

Project Timeline and Next Steps

Process Step	Timeline
Appeals Period	July 14 – August 15
Adjudication of Appeals	August 15 – September 12
Final Report	September 26

Questions?

NATIONAL QUALITY FORUM