
CBE ID

4440e

Title

Percent of hospitalized pneumonia patients with chest imaging confirmation

Project

Management of Acute Events, Chronic Disease, Surgery, and Behavioral Health

Endorsement Status

Endorsed

Is Under Review

No

Next Maintenance Cycle

Spring 2029

Previous Endorsement Cycle

Spring 2024

Initial Endorsement

Sun, 09/01/2024 - 10:05

Steward

University of Utah

1.0 New or Maintenance

New

1.3 Electronic Clinical Quality Measure (eCQM)

Yes

1.6 Measure Description

The chest imaging-confirmed measure of pneumonia diagnosis is a process measure of inpatient hospitalizations that identifies the proportion of adult patients hospitalized patients with a discharge diagnosis of pneumonia and who received systemic or oral antimicrobials at any time during admission who received chest imaging that supported the diagnosis of pneumonia, as recommended by clinical practice guidelines. The measure applies to a target population of adult hospitalized patients.

1.7 Composite Measure

No

1.7 Measure Type

Process

1.8 Level of Analysis

Facility

1.9 Care Setting

Hospital: Inpatient

1.10 Measure Rationale

As the leading infectious cause of death in the United States, the leading source of sepsis, and one of the most common reasons for adult hospitalizations, pneumonia is an important target for quality measurement and improvement. Prior measures of quality in pneumonia have driven important improvements for timely treatment and improved outcomes. However, past measures have also been challenged by difficulties in ensuring a consistent target population due to the subjectivity of diagnosis and the burden of manual chart review.

There are no gold standard criteria for the diagnosis of pneumonia, and signs and symptoms that are suggestive of pneumonia can overlap with other diagnoses. The presence of an infiltrate on chest imaging is the clinical sign that carries the most face validity as an objective verification of the diagnosis. Clinical practice guidelines strongly recommend verification of a pneumonia diagnosis with chest imaging. Confirmation of a diagnosis of pneumonia with chest imaging has been suggested as a performance metric by US professional societies and is listed as a practice standard in the United Kingdom. The availability of chest imaging is universal in US hospitals and is integrated into modern EHR's in searchable format to support clinical operations. Despite its availability, however, 10-30% of patients diagnosed with pneumonia lack positive chest imaging. With advances in the electronic medical record, an electronic Clinical Quality Measure that identifies chest imaging confirmation of diagnoses of pneumonia would 1) improve diagnostic accuracy in pneumonia, and 2) serve as an eCQM-based foundation for all quality measures in pneumonia.

The eCQM for diagnostic quality in pneumonia also balances with other current quality efforts, including reducing costs and improving 30-day mortality for pneumonia hospitalizations, reducing inappropriate antibiotics for bronchitis, and improving timely treatment and outcomes for sepsis. All of these measures could promote over-diagnosis of pneumonia, since patients without radiographic abnormalities are more likely to represent bronchitis, a disease with generally better outcomes, and for which clinicians are increasingly encouraged to avoid antibiotic use. Reporting the proposed eCQM in conjunction with the existing quality measures would promote meaningful quality improvement by ensuring accurate target populations and reporting.

1.12a Attach MADiE Output

[ChestImagingConfirmedPneumonia-v1-0-000-QDM-5-6.zip](#)

1.13 Data Dictionary

Not attached. I attest that all information will be provided where codes and/or value sets are needed (1.14a - 1.15c).

1.13a Attach Data Dictionary

[Value Sets and Data Dictionary.xlsx](#)

1.14 Numerator

Number of adult patients identified in the denominator (inpatient admission with a discharge diagnosis code at any position for pneumonia and administered systemic antimicrobials at any time during hospitalization) with evidence of a chest image consistent with pneumonia during their hospital admission or within 48 hours preceding the inpatient admission.

1.14a Numerator Details

A case is included in the numerator if it is in the denominator and has a chest image consistent with pneumonia during their hospital admission or within 48 hours preceding the inpatient admission (**Figure 1, please see Supplemental Materials attachment page 1**).

The list of codes and doi for value sets for imaging is provided in the data dictionary attached above in the sheet entitled "VSAC_ChestImagingforPneumoniaGrp". This list includes the CPT and HCPCS coding systems. Definitions of imaging consistent with pneumonia are provided in the "VSAC_AbnormalFindingChestImaging" spreadsheet. Chest imaging consistent with pneumonia can be identified in the EHR either as structured data using OID 2.16.840.1.113762.1.4.1264.21 and 2.16.840.1.113762.1.4.1264.12 (Sheets "VSAC_AbnormalFindingChestImaging" and "VSAC_ChestImagingForPneumoniaGrp" in the attachment or as unstructured data via natural language processing (NLP) of chest imaging reports.

1.15 Denominator

Number of inpatient admissions with a discharge diagnosis code at any position for pneumonia and administered systemic antimicrobials at any time during hospitalization or within 24 hours preceding admission.

1.15a Denominator Details

The denominator meets the following criteria:

1. An adult (age ≥ 18 years) at time of inpatient hospitalization
2. An index inpatient hospitalization (see attached value sets, Sheet Name: "VSAC_EncounterInpatient") with a discharge diagnosis of pneumonia at any position (i.e., the pneumonia diagnosis code may occur as a primary, principal, or secondary position). See attached data dictionary (Sheet Name: "VSAC_InpatientPneumoniaDiagnosis") for relevant ICD-9CM, ICD-10CM, and SNOMED-CT codes for pneumonia.
3. Receipt of systemic [oral or intravenous] antimicrobials (see attached value set for list of eligible antimicrobials and corresponding RxNorm CUI, Sheet Name:

“VSAC_SystemicAntimicrobials”) at any time during index hospitalization or in the 24 hours preceding hospital admission.

1.15b Denominator Exclusions

None

1.15c Denominator Exclusions Details

None

1.16 Type of Score

Rate/proportion

1.17 Measure Score Interpretation

Better performance = Higher score

1.18 Calculation of Measure Score

See **Figure 2** for diagram (attached in Fig2_Item1_18_Measure score calculation diagram.pdf).

Step 1: Define Target Population and Denominator:

For a one year period, identify all patients 18 and older with an inpatient hospitalization where the discharge diagnosis includes pneumonia (at any diagnosis coding position). Restrict to only patients who also received a systemic antimicrobial at any time during the index hospitalization or within 24 hours preceding admission. No other denominator exclusions.

Step 2: Define the Numerator:

From the denominator population, identify patients who received chest imaging (i.e., CT, x-ray) consistent with pneumonia during the index hospitalization or within 48 hours preceding admission.

Step 3: Calculate the Percent Concordance:

Divide the number of patients in the numerator with chest imaging concordant with discharge diagnosis of pneumonia (Step 2) by the number of patients in the denominator (Step 1) and multiply by 100. The measure is reported as a percentage (X% concordance). A higher percent concordance corresponds to better performance on the measure.

1.18a Attach measure score calculation diagram

[Fig2_Item1_18_Measure score calculation diagram.pdf](#)

1.19 Measure Stratification Details

The measure is not stratified.

1.20 Types of Data Sources

Electronic Health Records

1.25 Data Source Details

Electronic health record (EHR) data from inpatient hospital admissions, including discharge diagnosis codes, pharmacy and medication administration, and imaging records. These data are all collected routinely during usual clinical care through the process of inpatient hospitalizations.

Value sets defining these four concepts are available in the value sets (OIDs: 2.16.840.1.113762.1.4.1264.21; 2.16.840.1.113762.1.4.1264.1; 2.16.840.1.113762.1.4.1264.12; 2.16.840.1.113762.1.4.1264.8) and in the data dictionary attachment (Spreadsheet Names: "VSACAbnormalFindingChestImaging", "VSAC_ChestImagingforPneumoniaGrp", "VSACInpatientPneumoniaDiagnosis", "VSAC_SystemicAntimicrobials"). If a hospital system does not map abnormal chest imaging codes within their workflow, there are several tools either available or under development to extract evidence of pneumonia including structured reporting,¹ natural language processing from chest imaging reports^{2,3} and image processing directly applied to images.⁴⁻⁶ For this proposal, we developed and validated a rule-based natural language processing tool that has been made publicly available for use (see *Feasibility* and *Validity* sections below for details). The full list of references cited throughout the full submission responses are provided at the end of our response to item 2.2 *Evidence of Measure Importance*.

1.26 Minimum Sample Size

Table 1 indicates the minimum annual number of hospitalizations needed to meet criteria for the denominator to reach each target reliability level at a given facility. If a facility has fewer than the minimum number, the hospital is still encouraged to report performance on this measure. This approach is consistent with a previous inpatient community-acquired pneumonia measure set (PN-3a: Blood Cultures Performed within 24 Hours Prior to or 24 Hours After Hospital Arrival for Patients who were Transferred or Admitted to the ICU within 24 Hours of Hospital Arrival; PN-6: Initial Antibiotic Selection for Community-Acquired Pneumonia [CAP in Immunocompetent

Patients]).⁷ The minimum recommended sample sizes are smaller than those previously required (60 pneumonia patients annually for median-sized hospitals).⁷

Table 1. Required sample sizes for range of target reliability levels. Values are estimated by the Spearman Brown prophecy formula⁸⁻¹⁰ based upon the median number of patients meeting denominator criteria across 100 Veterans Affairs (VA) facilities in calendar year 2021 (n=70 patients) and the median accountable entity level reliability (0.617) were used. Reliability of a specific facility is a function of both sample size and measure performance. The provided reliability estimates assume the same performance score on the proposed measure across all facilities, with only sample size varying. Additional details on how the reliability estimate of 0.617 was calculated are provided in the Reliability section.

Target Reliability	Estimated Minimum Sample Size
0.4	29
0.5	43
0.6	65
0.7	101
0.8	174
0.9	391

2.1 Attach Logic Model

[Logic Model.pdf](#)

2.2 Evidence of Measure Importance

Pneumonia is a lower respiratory tract infection of the lung that causes inflammation in the alveoli, or airspaces.¹¹ It is a leading cause of sepsis,¹² hospitalizations,^{13,14} and death¹⁵ from infectious disease in the United States. Community-acquired pneumonia (CAP), or pneumonia acquired outside of healthcare settings, is responsible for 1.5 million annual hospitalizations in the US with an inpatient mortality of 6.5%.¹⁶ Hospital-acquired pneumonia (HAP) is far more rare than CAP (.5% versus 10% of all hospitalizations)^{16,17} but is one of the most common healthcare-associated infections accounting for an estimated 7% of all hospital deaths¹⁸ and with an individual mortality rate of over 25%.^{19,20} Annually, \$4.4-\$10 billion is spent on inpatient treatment of CAP, and inpatient treatment represents more than half of all treatment costs; each inpatient encounter costs nearly \$11,000.²¹⁻²³

Although pneumonia is common, the initial diagnosis can be uncertain and subjective. Pneumonia is one of the most common conditions associated with diagnostic error.²⁴ Without chest imaging – either through chest X-ray, computerized tomography (CT), or lung ultrasound – to confirm alveolar inflammation, pneumonia is difficult to clinically distinguish from acute bronchitis, for which antibiotics have minimal benefit and are not recommended.^{25,26} Practitioners consistently overestimate the probability of pneumonia in the absence of chest imaging results.²⁷ Clinical judgment in the absence of chest imaging demonstrates a sensitivity and specificity as low as 27%.^{28,29} While chest imaging is not perfect (chest radiography has lower sensitivity than CT), the use of chest imaging significantly increases diagnostic certainty of pneumonia and has been shown to improve the specificity and PPV of pneumonia diagnosis.²⁹ United States and international guidelines for both CAP and HAP require the use of chest imaging to diagnose pneumonia,³⁰⁻³⁴ with strong recommendations with moderate evidence and suggestions to consider the use of chest imaging as a quality measure in CAP.³⁵ In the United Kingdom, the National Institute for Healthcare Excellence has listed the use of chest imaging to support the diagnosis of pneumonia as a practice standard.³⁶

Despite existing clinical guidelines, evidence of the importance of chest imaging to diagnosis, and the universal availability of advanced chest imaging in emergency departments and hospitals, the diagnosis of pneumonia without a confirmatory chest image is still common in these settings, with approximately 10-30% of all CAP diagnoses made in the absence of positive imaging.^{23,37-39} Prior quality measures focused on timely treatment likely improved outcomes for pneumonia and sepsis, but some have suggested that they may have also led to more inappropriate diagnoses of CAP and antibiotic overuse.^{40,41} Overdiagnosis of pneumonia in the absence of chest imaging to support the diagnosis can lead to unnecessary antimicrobial use, unnecessary microbiological testing, unnecessary admission, and delays in diagnosing a patient's true condition.⁴⁰⁻⁴² For every day of unnecessary antibiotic treatment, patients inappropriately diagnosed with CAP experience a 5% increase inpatient-reported adverse events.⁴³

References (for full submission, not just 2.2 *Evidence of Measure Importance*):

1. Morgan TA, Helibrun ME, Kahn CE, Jr. Reporting initiative of the Radiological Society of North America: progress and new directions. *Radiology*. Dec 2014;273(3):642-5. doi:10.1148/radiol.14141227
2. Elkin PL, Froehling D, Wahner-Roedler D, et al. NLP-based identification of pneumonia cases from free-text radiological reports. *AMIA Annu Symp Proc*. Nov 6 2008;2008:172-6.
3. Dublin S, Baldwin E, Walker RL, et al. Natural Language Processing to identify pneumonia from radiology reports. *Pharmacoepidemiol Drug Saf*. Aug 2013;22(8):834-41. doi:10.1002/pds.3418
4. Dean N, Irvin JA, Samir PS, Jephson A, Conner K, Lungren MP. Real-time electronic interpretation of digital chest images using artificial intelligence in emergency department patients suspected of pneumonia. *Eur Respir J*. 2019;54(suppl 63):OA3309.

- doi:10.1183/13993003.congress-2019.OA3309
5. Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJ. Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*. 2019;145:511-518.
 6. Shih G, Wu CC, Halabi SS, et al. Augmenting the National Institutes of Health Chest Radiograph Dataset with Expert Annotations of Possible Pneumonia. *Radiol Artif Intell*. 2019;1(1):e180041. doi:10.1148/ryai.2019180041
 7. The Joint Commission. Specifications manual for Joint Commission National Quality Measures (v2015A): Pneumonia (PN). 2014. Available at: https://manual.jointcommission.org/releases/TJC2015A/Pneumonia.html#A_4.... Accessed 15 Feb 2024.
 8. van Ast JF, Talmon JL, Renier WO, Hasman A. An approach to knowledge base construction based on expert opinions. *Methods Inf Med*. 2004;43(4):427-32.
 9. Zimmerman DW, Williams RH, Burkheimer GJ. Dependence of reliability of multiple-choice tests upon number of choices per item: prediction from the Spearman-Brown formula. *Psychol Rep*. Dec 1966;19(3):1239-43. doi:10.2466/pr0.1966.19.3f.1239
 10. de Vet HCW, Mokkink LB, Mosmuller DG, Terwee CB. Spearman-Brown prophecy formula and Cronbach's alpha: different faces of reliability and opportunities for new applications. *J Clin Epidemiol*. May 2017;85:45-49. doi:10.1016/j.jclinepi.2017.01.013
 11. Mackenzie G. The definition and classification of pneumonia. *Pneumonia*. 2016;8:14. doi:10.1186/s41479-016-0012-z
 12. Novosad SA, Sapiano MR, Grigg C, et al. Vital Signs: Epidemiology of Sepsis: Prevalence of Health Care Factors and Opportunities for Prevention. *MMWR Morb Mortal Wkly Rep*. Aug 26 2016;65(33):864-9. doi:10.15585/mmwr.mm6533e1
 13. McDermott K, Roemer M. Most frequent principal diagnoses for inpatient stays in U.S. hospitals, 2018. (2021). In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006. Statistical Brief #277.
 14. Hayes BH, Haberling DL, Kennedy JL, Varma JK, Fry AM, Vora NM. Burden of Pneumonia-Associated Hospitalizations: United States, 2001-2014. *Chest*. Feb 2018;153(2):427-437. doi:10.1016/j.chest.2017.09.041
 15. Heron M. Deaths: Leading Causes for 2018. *Natl Vital Stat Rep*. May 2021;70(4):1-115.
 16. Ramirez JA, Wiemken TL, Peyrani P, et al. Adults Hospitalized With Pneumonia in the United States: Incidence, Epidemiology, and Mortality. *Clin Infect Dis*. Nov 13 2017;65(11):1806-1812. doi:10.1093/cid/cix647
 17. Magill SS, Edwards JR, Bamberg W, et al. Multistate point-prevalence survey of health care-associated infections. *N Engl J Med*. Mar 27 2014;370(13):1198-208. doi:10.1056/NEJMoa1306801
 18. Jones BE, Sarvet AL, Ying J, et al. Incidence and Outcomes of Non-Ventilator-Associated Hospital-Acquired Pneumonia in 284 US Hospitals Using Electronic Surveillance Criteria. *JAMA Network Open*. 2023;6(5):e2314185-e2314185. doi:10.1001/jamanetworkopen.2023.14185
 19. Jones BE, Sarvet AL, Ying J, et al. Incidence and Outcomes of Non-Ventilator-Associated Hospital-Acquired Pneumonia in 284 US Hospitals Using Electronic Surveillance Criteria. *JAMA Netw Open*. May 01 2023;6(5):e2314185. doi:10.1001/jamanetworkopen.2023.14185
 20. Papazian L, Klompas M, Luyt CE. Ventilator-associated pneumonia in adults: a narrative review. *Intensive Care Med*. May 2020;46(5):888-906. doi:10.1007/s00134-020-05980-0
 21. Niederman MS, McCombs JS, Unger AN, Kumar A, Popovian R. The cost of treating

- community-acquired pneumonia. *Clin Ther.* Jul-Aug 1998;20(4):820-37.
doi:10.1016/s0149-2918(98)80144-6
22. Tong S, Amand C, Kieffer A, Kyaw MH. Trends in healthcare utilization and costs associated with pneumonia in the United States during 2008-2014. *BMC Health Serv Res.* Sep 14 2018;18(1):715. doi:10.1186/s12913-018-3529-4
23. Jain S, Self WH, Wunderink RG, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N Engl J Med.* Jul 30 2015;373(5):415-27.
doi:10.1056/NEJMoa1500245
24. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med.* Mar 25 2013;173(6):418-25.
doi:10.1001/jamainternmed.2013.2777
25. Smucny J, Fahey T, Becker L, Glazier R. Antibiotics for acute bronchitis. The Cochrane database of systematic reviews. Oct 18 2004;(4):Cd000245.
doi:10.1002/14651858.CD000245.pub2
26. Braman SS. Chronic Cough Due to Acute Bronchitis: ACCP Evidence-Based Clinical Practice Guidelines. *Chest.* 2006/01/01/ 2006;129(1, Supplement):95S-103S.
doi:https://doi.org/10.1378/chest.129.1_suppl.95S
27. Morgan DJ, Pineles L, Owczarzak J, et al. Accuracy of Practitioner Estimates of Probability of Diagnosis Before and After Testing. *JAMA Intern Med.* Jun 01 2021;181(6):747-755.
doi:10.1001/jamainternmed.2021.0269
28. Wootton D, Feldman C. The diagnosis of pneumonia requires a chest radiograph (x-ray)-yes, no or sometimes? *Pneumonia.* 2014;5(Suppl 1):1-7. doi:10.15172/pneu.2014.5/464
29. Klompas M. Clinical evaluation and diagnostic testing for community-acquired pneumonia in adults. In: UpToDate, Wolters Kluwer; 2023, Eds., Ramirez J, Bond S, Dieffenbach P.
30. Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis.* Mar 01 2007;44 Suppl 2(Suppl 2):S27-72.
doi:10.1086/511159
31. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med.* Oct 1 2019;200(7):e45-e67. doi:10.1164/rccm.201908-1581ST
32. File TM, Ramirez JA. Community-Acquired Pneumonia. *N Engl J Med.* Aug 17 2023;389(7):632-641. doi:10.1056/NEJMcp2303286
33. National Institute for Health and Care Excellence (NICE), London. (2023). Pneumonia in adults: diagnosis and management. NICE Guideline, No 191. ISBN-13: 978-1-4731-5518-3.
34. Kalil AC, Metersky ML, Klompas M, et al. Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis.* Sep 01 2016;63(5):e61-e111. doi:10.1093/cid/ciw353
35. Bartlett JG, Dowell SF, Mandell LA, File TM, Musher DM, Fine MJ. Practice guidelines for the management of community-acquired pneumonia in adults. Infectious Diseases Society of America. *Clin Infect Dis.* Aug 2000;31(2):347-82. doi:10.1086/313954
36. National Institute for Health and Care Excellence (NICE). Pneumonia in adults, quality standard [QS110]: Quality statement 3-- Chest x-ray and diagnosis within 4 hours of hospital presentation. <https://www.nice.org.uk/guidance/qs110/chapter/quality-statement-3-ches...>
37. Chandra A, Nicks B, Maniago E, Nouh A, Limkakeng A. A multicenter analysis of the ED

- diagnosis of pneumonia. *Am J Emerg Med*. Oct 2010;28(8):862-5. doi:10.1016/j.ajem.2009.04.014
38. Marrie TJ, Huang JQ. Low-risk patients admitted with community-acquired pneumonia. *Am J Med*. Dec 2005;118(12):1357-63. doi:10.1016/j.amjmed.2005.06.035
39. Atamna A, Shiber S, Yassin M, Drescher MJ, Bishara J. The accuracy of a diagnosis of pneumonia in the emergency department. *Int J Infect Dis*. Dec 2019;89:62-65. doi:10.1016/j.ijid.2019.08.027
40. Kanwar M, Brar N, Khatib R, Fakhri MG. Misdiagnosis of community-acquired pneumonia and inappropriate utilization of antibiotics: side effects of the 4-h antibiotic administration rule. *Chest*. Jun 2007;131(6):1865-9. doi:10.1378/chest.07-0164
41. Welker JA, Huston M, McCue JD. Antibiotic timing and errors in diagnosing pneumonia. *Arch Intern Med*. Feb 25 2008;168(4):351-6. doi:10.1001/archinternmed.2007.84
42. Ang CS, Kelvin Beh KM, Yeang LJ, et al. Misdiagnosis of community-acquired pneumonia in patients admitted to respiratory wards, Penang General Hospital. *Med J Malaysia*. Jul 2020;75(4):385-390.
43. Gupta AB, Flanders SA, Petty LA, et al. Inappropriate Diagnosis of Pneumonia Among Hospitalized Adults. *JAMA Intern Med*. Mar 25 2024;doi:10.1001/jamainternmed.2024.0077
44. Shorr AF, Owens RC. Guidelines and quality for community-acquired pneumonia: measures from the Joint Commission and the Centers for Medicare and Medicaid Services. *Am J Health Syst Pharm*. Jun 15 2009;66(12 Suppl 4):S2-7. doi:10.2146/090087a
45. The Joint Commission. Specifications manual for Joint Commission National Quality Core Measures (2010A1): pneumonia (PN). 2010. Available at: <https://manual.jointcommission.org/releases/archive/TJC2010B/Pneumonia....> Accessed 22 Feb 2024.
46. Munro SC, Baker D, Giuliano KK, et al. Nonventilator hospital-acquired pneumonia: A call to action. *Infect Control Hosp Epidemiol*. Aug 2021;42(8):991-996. doi:10.1017/ice.2021.239
47. Self WH, Courtney DM, McNaughton CD, Wunderink RG, Kline JA. High discordance of chest x-ray and computed tomography for detection of pulmonary opacities in ED patients: implications for diagnosing pneumonia. *Am J Emerg Med*. Feb 2013;31(2):401-5. doi:10.1016/j.ajem.2012.08.041
48. Bourcier JE, Paquet J, Seinger M, et al. Performance comparison of lung ultrasound and chest x-ray for the diagnosis of pneumonia in the ED. *Am J Emerg Med*. Feb 2014;32(2):115-8. doi:10.1016/j.ajem.2013.10.003
49. Centers for Medicare and Medicaid. Analysis of Topped-Out Measures Finalized for the PY 2016 ESRD QIP. (2014). <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Ins....> Accessed 09 Nov, 2023.
50. Centers for Medicare and Medicaid. Merit-Based Incentive Payment System (MIPS): simple pneumonia with hospitalization measure, 2020 performance period. 2019. Available at: https://qpp.cms.gov/docs/cost_specifications/2019-12-17-mif-ebcm-pna-ho.... Accessed 23 Feb 2024.
51. Centers for Medicare and Medicaid. Quality ID#111: Pneumococcal vaccination status for older adults. 2023. Available at: https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measure.... Accessed 23 Feb 2024.
52. Centers for Medicare and Medicaid. Eligible Hospital / Critical Access Hospital eQMs: hospital quality reporting table of eQMs. Nov

2023. <https://ecqi.healthit.gov/sites/default/files/Hybrid-EH-CAH-MeasuresTab...>. Accessed 23 Feb 2024.
53. Agency for Healthcare Research and Quality. AHRQ IQI technical documentation, version v2023. Rockville, MD. 2023. Available at: https://qualityindicators.ahrq.gov/measures/iqi_resources. Accessed 02 Apr 2024.
54. Fee C, Weber EJ. Identification of 90% of patients ultimately diagnosed with community-acquired pneumonia within four hours of emergency department arrival may not be feasible. *Ann Emerg Med*. May 2007;49(5):553-9. doi:10.1016/j.annemergmed.2006.11.008
55. Metersky ML, Ma A, Bratzler DW, Houck PM. Predicting bacteremia in patients with community-acquired pneumonia. *Am J Respir Crit Care Med*. Feb 1 2004;169(3):342-7.
56. Partnership for Quality Measurement. Inappropriate diagnosis of community-acquired pneumonia (CAP) in hospitalized medical patients; Abbreviated form: in appropriate diagnosis of CAP -- CBE ID 3671. 2022. Available at: <https://p4qm.org/measures/3671>. Accessed 23 Feb 2024.
57. Rothberg MB, Pekow PS, Priya A, Lindenauer PK. Variation in diagnostic coding of patients with pneumonia and its association with hospital risk-standardized mortality rates: a cross-sectional analysis. *Ann Intern Med*. Mar 18 2014;160(6):380-8. doi:10.7326/M13-1419
58. Lindenauer PK, Lagu T, Shieh MS, Pekow PS, Rothberg MB. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. *JAMA*. Apr 4 2012;307(13):1405-13. doi:10.1001/jama.2012.384
59. Ruhnke GW, Coca-Perrillon M, Kitch BT, Cutler DM. Trends in mortality and medical spending in patients hospitalized for community-acquired pneumonia: 1993-2005. *Med Care*. Dec 2010;48(12):1111-6. doi:10.1097/MLR.0b013e3181f38006
60. Watkins RR, Lemonovich TL. Diagnosis and management of community-acquired pneumonia in adults. *Am Fam Physician*. Jun 1 2011;83(11):1299-306.
61. Woodhead M, Blasi F, Ewig S, et al. Guidelines for the management of adult lower respiratory tract infections--full version. *Clin Microbiol Infect*. Nov 2011;17 Suppl 6(Suppl 6):E1-59. doi:10.1111/j.1469-0691.2011.03672.x
62. National Institute for Health and Care Excellence (NICE). Pneumonia (including community acquired pneumonia). 2016. <http://www.nice.org.uk/guidance/cg191>
63. Bradley JS, Byington CL, Shah SS, et al. The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. *Clin Infect Dis*. Oct 2011;53(7):e25-76. doi:10.1093/cid/cir531
64. Sacco AY, Self QR, Worswick EL, et al. Patients' Perspectives of Diagnostic Error: A Qualitative Study. *J Patient Saf*. Dec 1 2021;17(8):e1759-e1764. doi:10.1097/pts.0000000000000642
65. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: Human-centered design of the patient journey for patient safety. *Appl Ergon*. Apr 2020;84:103033. doi:10.1016/j.apergo.2019.103033
66. Gupta AB, Flanders SA, Petty LA, et al. Inappropriate Diagnosis of Pneumonia Among Hospitalized Adults. *JAMA Internal Medicine*. 2024;doi:10.1001/jamainternmed.2024.0077
67. Irvin JA, Pareek A, Long J, et al. CheXED: Comparison of a Deep Learning Model to a Clinical Decision Support System for Pneumonia in the Emergency Department. *J Thorac Imaging*. May 1 2022;37(3):162-167. doi:10.1097/rti.0000000000000622
68. Chapman AB, Peterson KS, Rutter E, et al. Development and evaluation of an interoperable natural language processing system for identifying pneumonia across clinical settings of

- care and institutions. *JAMIA Open*. Dec 2022;5(4):ooac114. doi:10.1093/jamiaopen/ooac114
69. National Library of Medicine. Unified Medical Language System (UMLS): RxNORM. Accessed 22 Jan 2024. Available at: <https://www.nlm.nih.gov/research/umls/rxnorm/index.html>
70. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276-82.
71. Adams JL. The reliability of provider profiling: a tutorial. Santa Monica, CA: RAND Corporation; 2009. Available at: https://www.rand.org/pubs/technical_reports/TR653.html.
72. Morris C. Parametric empirical Bayes inference: theory and applications. *J Am Stat Assoc*. 1983;78(381):47.
73. Jones BE, Haroldsen C, Madaras-Kelly K, et al. In Data We Trust? Comparison of Electronic Versus Manual Abstraction of Antimicrobial Prescribing Quality Metrics for Hospitalized Veterans With Pneumonia. *Med Care*. Jul 2018;56(7):626-633. doi:10.1097/MLR.0000000000000916
74. Loeb MB, Carusone SBC, Marrie TJ, et al. Interobserver Reliability of Radiologists' Interpretations of Mobile Chest Radiographs for Nursing Home-Acquired Pneumonia. *J Am Med Dir Assoc*. 2006/09/01/ 2006;7(7):416-419. doi:<https://doi.org/10.1016/j.jamda.2006.02.004>
75. Hripcsak G, Kuperman GJ, Friedman C, Heitjan DF. A reliability study for evaluating information extraction from radiology reports. *J Am Med Inform Assoc*. Mar-Apr 1999;6(2):143-50. doi:10.1136/jamia.1999.0060143
76. Ramirez JA, Wiemken TL, Peyrani P, et al. Adults Hospitalized With Pneumonia in the United States: Incidence, Epidemiology, and Mortality. *Clin Infect Dis*. 2017;65(11):1806-1812. doi:10.1093/cid/cix647
77. Burton DC, Flannery B, Bennett NM, et al. Socioeconomic and racial/ethnic disparities in the incidence of bacteremic pneumonia among US adults. *Am J Public Health*. Oct 2010;100(10):1904-11. doi:10.2105/ajph.2009.181313
78. Altawalbeh SM, Wateska AR, Nowalk MP, et al. Societal Cost of Racial Pneumococcal Disease Disparities in US Adults Aged 50 Years or Older. *Appl Health Econ Health Policy*. 2024/01/01 2024;22(1):61-71. doi:10.1007/s40258-023-00854-0
79. Nowalk MP, Wateska AR, Lin CJ, et al. Racial Disparities in Adult Pneumococcal Vaccination Indications and Pneumococcal Hospitalizations in the U.S. *J Natl Med Assoc*. Oct 2019;111(5):540-545. doi:10.1016/j.jnma.2019.04.011
80. Hausmann LRM, Ibrahim SA, Mehrotra A, et al. Racial and Ethnic Disparities in Pneumonia Treatment and Mortality. *Med Care*. 2009;47(9):1009-1017.
81. Wiemken TL, Carrico RM, Furmanek SP, et al. Socioeconomic Position and the Incidence, Severity, and Clinical Outcomes of Hospitalized Patients With Community-Acquired Pneumonia. *Public Health Rep*. May/Jun 2020;135(3):364-371. doi:10.1177/0033354920912717
82. Downing NS, Wang C, Gupta A, et al. Association of Racial and Socioeconomic Disparities With Outcomes Among Patients Hospitalized With Acute Myocardial Infarction, Heart Failure, and Pneumonia: An Analysis of Within- and Between-Hospital Variation. *JAMA Network Open*. 2018;1(5):e182044-e182044. doi:10.1001/jamanetworkopen.2018.2044
83. Prendki V, Scheffler M, Huttner B, et al. Low-dose computed tomography for the diagnosis of pneumonia in elderly patients: a prospective, interventional cohort study. *Eur Respir J*. May 2018;51(5)doi:10.1183/13993003.02375-2017
84. Trent S, Havranek E, Ginde A, Haukoos J. 291EMF Effect of Audit and Feedback on

Physician Adherence to Clinical Practice Guidelines for Pneumonia and Sepsis. *Ann Emerg Med*. 2016/10/01/ 2016;68(4, Supplement):S114.

doi:<https://doi.org/10.1016/j.annemergmed.2016.08.306>

85. Dean NC, Vines CG, Carr JR, et al. A Pragmatic, Stepped-Wedge, Cluster-controlled Clinical Trial of Real-Time Pneumonia Clinical Decision Support. *Am J Respir Crit Care Med*. Jun 1 2022;205(11):1330-1336. doi:10.1164/rccm.202109-2092OC
86. Kukhareva P, Weir CR, Staes C, Borbolla D, Slager S, Kawamoto K. Integration of Clinical Decision Support and Electronic Clinical Quality Measurement: Domain Expert Insights and Implications for Future Direction. *AMIA Annu Symp Proc*. 2018;2018:700-709.
87. Wyatt R. Reducing Hospital-Acquired Pressure Injuries through Measure-vention. *Adv Skin Wound Care*. Jan 1 2022;35(1):43-47. doi:10.1097/01.Asw.0000801528.12103.88
88. Dean NC, Jones BE, Jones JP, et al. Impact of an Electronic Clinical Decision Support Tool for Emergency Department Patients With Pneumonia. *Ann Emerg Med*. Nov 2015;66(5):511-20. doi:10.1016/j.annemergmed.2015.02.003
89. Dean NC, Jones BE, Ferraro JP, Vines CG, Haug PJ. Performance and utilization of an emergency department electronic screening tool for pneumonia. *JAMA Intern Med*. Apr 22 2013;173(8):699-701. doi:10.1001/jamainternmed.2013.3299

2.3 Anticipated Impact

Failure to obtain confirmatory chest imaging to diagnose pneumonia is common^{23,37-39} and can lead to antibiotic and resource overuse, increased antibiotic resistance, and delayed treatment of the true condition with accompanying downstream adverse events.⁴⁰⁻⁴² We expect direct effects of this measure on both antibiotic use and outcomes. Treatment of patients over-diagnosed with CAP has been associated with a 5% increase in patient-reported adverse effects of antibiotics for every day of unnecessary antibiotics.⁴³ In empiric testing of the association between the measure and patient outcomes across VA hospitals, we estimated that patients diagnosed with pneumonia but with a negative chest image would have a 44% absolute decrease in the risk of receiving empiric antibiotics than those without an inappropriate diagnosis, after adjusting for baseline patient comorbidities and clinical illness severity. We also observed that facilities with greater alignment between diagnosis and chest imaging had a greater likelihood of obtaining CT scans and lower 30-day readmissions among all hospitalizations (See *Validity* section). We thus expect to see an increase in appropriate use of imaging to verify pneumonia diagnoses, measurable reduction in inappropriate antibiotic use and adverse events due to antibiotics at the hospital level, and a possible reduction of hospital readmission at the system level.

If the currently proposed measure is implemented as a publicly reported measure tied to reimbursement, we expect a fairly rapid improvement of the measure performance through 3 mechanisms:

1. improvement of collection, integration, and documentation of chest imaging data within and between EHR's;
2. reduction of inappropriate use of the diagnostic label of pneumonia when it is not supported by clinical data; and
3. increase in appropriate additional workup to confirm pneumonia diagnosis for patients with

high suspicion but initial negative workup.

We anticipate that the measure may eventually no longer be needed as a quality measure itself. However, we expect this eCQM to contribute to additional eCQMs by identifying a standardized target population of image-confirmed pneumonia. Both CAP and HAP have historically been important targets of quality improvement or safety efforts. For CAP, prior and existing clinical quality measures have focused on process and outcome measures such as time to antimicrobial therapy, appropriate microbiological testing, antimicrobial prescribing and treatment, secondary interventions such as vaccination or tobacco cessation, and mortality and readmission outcomes.^{7,44,45} Prior inpatient CAP measures have begun by excluding patients who lack confirmatory chest imaging from the initial population.³⁸ This process has previously been achieved through manual chart abstraction, which is laborious and inconsistent. The proposed inpatient pneumonia eCQM uses data from the electronic medical record to obtain an initial target population, from which additional exclusions can be applied. This supports the transition of existing measures to eCQMs by promoting consistent definitions across inpatient pneumonia measures, allowing more valid performance comparisons and a shared understanding of the state of safety and quality issues in pneumonia.

Although HAP is much rarer than CAP, the proposed eCQM may also reduce HAP incidence by improving surveillance and prevention efforts. For both ventilator-associated and non-ventilator-associated HAP, a major barrier to prevention has been the inability to objectively track incidence due to the subjectivity of diagnosis codes, an issue which recently drove a joint task force of healthcare systems, public health agencies, and professional societies to call for improved HAP surveillance strategies.⁴⁶ As with treatment and other downstream care quality measures for CAP, adding consistency to the definition of pneumonia provided by the proposed eCQM provides a standard foundation from which surveillance and quality measures for HAP may also be used.

We anticipate several unintended consequences. Chest radiograph (X-ray) is the cheapest and most common imaging modality for the diagnosis of pneumonia, but its sensitivity is poor relative to CT.⁴⁷ Currently, patients with a high clinical suspicion of pneumonia but negative chest imaging, or identification of pneumonia on non-chest imaging, might be treated without additional imaging. Implementation of a measure that requires imaging confirmation of the diagnosis of pneumonia could drive an increase in the use of chest CT to improve diagnostic accuracy, which could increase costs and burden to patients and institutions. It may in rare cases be more appropriate to forego confirmatory imaging rather than pursue a more accurate imaging approach, as might be the case for very unstable patients or those with transmissible diseases. The increased cost and burden of CT's must be balanced by the benefit of advanced imaging in patients who may actually have a different diagnosis (such as pulmonary embolism, alternative infection, or malignancy) that can be unearthed. When this issue was discussed with our patient representatives in our technical expert panel (TEP), they described the process of obtaining a CT from their perspective and expressed a preference for obtaining this procedure over having an unconfirmed diagnosis. Full description of methods from the TEP are included in section 2.6

(Meaningfulness to Target Population) below.

Quotes from Patient TEP Participants (discussion context: obtaining a CT if chest X-ray is negative):

Patient 1: *"I would prefer not to endure every test under the sun but sometimes it's in my best interest for*

them to get the correct diagnosis so that they can treat me with the best medications that get me home."

Patient 2: *"I would rather have the test to be able to diagnose than what I went through and to get the*

initial diagnosis because I couldn't get them to test at all."

Measure implementation could drive greater use of point-of-care lung ultrasound (US), an emerging technology which is not currently recommended for confirming a pneumonia diagnosis but has higher sensitivity than chest X-ray and has been suggested as a less costly, more convenient, and potentially safer alternative to CT (for unstable patients) when chest X-rays are negative.⁴⁸ One major concern regarding lung US is the current lack of standards for equipment, training, image capture and archival, and documentation into the medical record for quality assurance. If lung US becomes a more common way to confirm a diagnosis of pneumonia, the proposed eCQM would be adapted to include data sources to capture this imaging data. Requiring electronic confirmation of positive lung US in the EHR could serve an important role in ensuring that the lung US is subject to the same reporting standards as current radiologic procedures.

We also anticipate potential changes to radiology workflow, including reporting (such as an increase in structured or more standardized reporting or more "hedging" language), an increase in dialogue between clinicians and radiologists (and potential for editing if clinical suspicion of pneumonia is very high but initial radiographic interpretation is negative), or an increase in the use of natural language processing or image processing software ("AI" applied directly to images) to support radiology reporting, documentation, or measure performance. We have tested the feasibility and validity of natural language processing to support this measure, but this is only one potential tool. The proposed eCQM is designed to be sufficiently flexible to adapt to these new approaches to radiology reporting.

2.4 Performance Gap

Evidence of Performance Gap

To examine variation and evaluate the degree we might expect a performance gap across hospitals, we used the VA healthcare system dataset to conduct interfacility comparisons of performance on the proposed eCQM. The initial population dataset of 103 VA acute care facilities and 89,767 hospitalizations diagnosed and treated for pneumonia covering admissions from 2015-2022 was restricted to a single year of the most recent data (2021) to match the one-year reporting period specified for this proposed eCQM. The most recent year of data included 100 VA hospitals and 8,253 hospitalizations with a discharge diagnosis of pneumonia where the patients also received antimicrobials (measure denominator). Facilities were sorted into deciles by eCQM performance score (**Table 2 , attachment 2_4a_PerformanceGapResults_CBE4440e.pdf**). A smoothed estimated of facility performance on the proposed eCQM that accounts for the effects of small denominators at smaller hospitals was also calculated using an empirical Bayes approach (**Table 3, attachment 2_4a_PerformanceGapResults_CBE4440e.pdf**). Results are presented in the *2.4a Performance Gap Results* attachment. We recognize that this is a new proposed eCQM measure, the VA may not be representative of other settings, and if endorsed, the performance gap would have to be evaluated in other settings.

Assessment of Measure being “Topped Out”

Using the same dataset as outlined for the performance gap analysis, we also analyzed whether there was evidence of the measure being “topped out” with insufficient variation on measure performance between reporting entities (**Table 4, attachment 2_4a_PerformanceGapResults_CBE4440e.pdf**). A measure is considered “topped out” if meeting two criteria: 1) the 75th and 90th percentile scores are statistically indistinguishable by two standard errors AND 2) the truncated coefficient of variation (TCV) was ≤ 0.10 .⁴⁹ We obtained the TCV by removing the upper and lower five percentiles of measure scores from the dataset and divided the standard deviation of this restricted dataset by the mean of the restricted dataset.

2.4a Attach Performance Gap Results

[2_4a_PerformanceGapResults_CBE4440e.pdf](#)

2.5 Health Care Quality Landscape

Pneumonia is a well-established target of clinical quality and safety monitoring. Table 5 (see *Supplemental Materials* attachment) lists existing and prior measures for pneumonia as well as existing measures for related conditions that may influence or be influenced by measures of pneumonia, including sepsis, COPD, and bronchitis.^{7,44,45,50-53}

Most prior and existing pneumonia performance measures have focused on quality of processes or outcomes in CAP once a diagnosis is made. Diagnosis quality has not been a focus of previous measures,⁴⁵ and concerns have been raised that the focus on timely tests and treatment, such as

prior antibiotic timing and blood culturing measures in pneumonia or current culturing and antibiotic measures or the current sepsis management bundle, may result in the overdiagnosis and overtreatment.^{40,54,55} Further, because CAP, COPD, and bronchitis can all present similarly and are indistinguishable from each other without chest imaging, attention to clinical outcomes for one diagnosis that may be subject to variation and shifts in diagnostic coding could have externalities to other diagnoses. For example, the avoidance of antibiotics for bronchitis could inadvertently increase the inappropriate diagnosis of pneumonia as a similar condition that justifies the use of antibiotics. Similarly, the existing outcome measures for pneumonia could inadvertently encourage misclassification of bronchitis, which has a lower mortality risk than pneumonia, resulting in an overall appearance of improved outcomes for pneumonia without actually improving outcomes. Without evaluation of diagnostic quality, measurement can suffer from difficulty in validating the target populations for all of these conditions.

No eQMs for pneumonia currently exist.⁵² Prior chart-review based pneumonia measures began by excluding patients who lacked confirmatory chest imaging from the initial population specification.^{7,45} One recently endorsed measure identifying inappropriate diagnosis of pneumonia (CBE 3671 – University of Michigan) includes chest imaging within a set of diagnostic criteria, but this measure requires manual chart abstraction.⁵⁶ We have collaborated with the developers of this measure to explore the feasibility and accuracy of supporting the chest imaging component with the proposed eQM.

Existing quality measures have also attempted to identify target populations focused on either community-acquired or healthcare-associated pneumonia, using diagnostic coding positions or present-on-admission criteria. These approaches have not been resilient to temporal or geographic coding variation.⁵⁷⁻⁵⁹ While the vast majority (over 95%) of pneumonia diagnoses are CAP, the currently proposed eQM denominator can be applied to either pneumonia type and can thus be used to identify a standard target population that can be further refined to capture CAP or HAP for other important measures.

The insufficiency of existing measures was also discussed by members of the Technical Expert Panel (TEP), along with the potential for the proposed eQM to support the development of new measures that address these insufficiencies:

*“Lots of patients are now categorized as sepsis, principal diagnosis and if that's done at different rates in different hospitals, it becomes very difficult to compare the outcomes for hospital A and hospital B. **So consistency of definition will have innumerable benefits to both the public, the people...receiving the care, as well as the entities that are measuring and reporting quality.**”* -- TEP participant 1 (clinician and pneumonia quality expert)

"I think there, you know, is feedback from clinicians that while you're using codes and it's not based on symptoms or chest X-ray and what not and then that's a gap. So I think that you pursuing this is really helpful and, and can benefit...the quality measure for CAP and in general." - TEP participant 2 (CDC representative)

"If we cannot trust in the DRG... and we have all these electronic medical records, how do we really know how many patients with pneumonia really die of pneumonia in a particular hospital if we don't even know who has pneumonia?" -- TEP participant 3 (clinician, pneumonia practice guideline author)

2.6 Meaningfulness to Target Population

We explored the value of the diagnostic quality in pneumonia for two major stakeholders – clinicians and patients - through two approaches: 1) literature review and 2) engagement with patients and clinicians in technical expert panels.

Literature review:

Pubmed, GoogleScholar, and guideline repositories were searched using the keywords "pneumonia", "diagnostic quality", "chest X-ray/CT/imaging", "patient perspective" and "clinician perspective."

Clinicians' value of diagnostic quality in pneumonia is most strongly demonstrated by the central role that professional society practice guidelines have placed on chest imaging diagnosis to standard management. The most recent clinical practice guidelines for community-acquired pneumonia from American Thoracic Society and Infectious Diseases Society of America (ATS/IDSA), the professional societies that represent pulmonary & critical care physicians and infectious disease physicians but that also included internists and emergency medicine providers in their committee, recommended the use of chest imaging to confirm a diagnosis of pneumonia.³¹ Prior IDSA practice guidelines rated chest imaging as a strong recommendation and also called for a performance indicator.³⁵ While expert guidelines published from American Academy of Family Physicians,⁶¹ Europe,⁶² and the United Kingdom,⁶³ also emphasize the importance of chest imaging where settings allow, pediatric guidelines recommend only chest imaging for hospitalized children,⁶⁴ and AHRQ states that not obtaining a chest image is reasonable if the diagnosis is certain and the patient does not have substantial dyspnea or hypoxemia. Hospitalized patients do not typically fit these criteria.

Patient perspectives on diagnostic quality in pneumonia are less well documented in the clinical literature. However, studies on patient experiences in general demonstrate that patients identify diagnostic error as a common experience, particularly in the hospital, with important consequences including emotional distress, health outcomes, and impairments on function.⁶⁵

Technical Expert Panel (TEP)

Full details surrounding the TEP methods are described in the supplemental attachment. In brief, meetings were conducted via videoconferencing (four during the initial measure development phase and an additional TEP during final measure specification of proposed eCQM (Table 6, pg. 3 *Supplemental Materials* attachment). For all five panels, an agenda with specific sets of questions tailored to the measure development phase were provided to panelists prior to the discussion session.

Illustrative quotes from the TEP are listed below. Patients from the TEP reported experiences of delayed or missed diagnoses as a result of failure to pursue imaging (quote 1), issues with diagnosis rather than treatment (quote 2), and the value of imaging to confirm infection, particularly in ambiguous presentations with nonspecific signs or symptoms (quote 3). Clinicians from the TEP reported value of diagnostic quality in pneumonia as 1) part of a broader effort to promote diagnostic quality (quote 5), as an important component to guideline-based pneumonia care (quote 6) and quality improvement (quote 7), and to antibiotic stewardship (quote 8).

1. [Patient representative; theme: importance of chest imaging to workup of dyspnea; quality gap] -

"When I turned 40, I was having a problem breathing...I spent three years going to probably 15 or 20

*different doctors. **Not one of them performed a chest x-ray** Finally an internist...said, well let's do an x-ray or, or a CT scan. And she did the x-ray and within 24 hours told me [my diagnosis]...*

2. [Patient representative; theme: quality gap] -

*"I've had pneumonia several times... when it comes to lung problems it's always been very good care, **and so more [issues] on the diagnostic side, less on the pneumonia [treatment] side.**"*

3. [Patient representative; theme: importance of chest imaging to workup of lower respiratory tract infections] -

"I never spike a fever.... Also, my white blood cell counts usually go really low. I'm sick.... [but providers will] say 'oh you don't have an infection' and then they'll take a

chest X-ray, and they go. 'Ohh my God.'
like, 'yeah, I know.'"

And they say you have a problem and it's

4. [Clinician representative; theme: face validity] -

"I, I think we need to remember what is the group of people that you're looking at... If you put a patient in the hospital for pneumonia, [and] you don't have an x-ray. This is kind of bad."

5. [Clinician & clinical practice guideline author; theme: value to standard pneumonia care] -

"We always say that these guidelines are for patient with pneumonia, but we don't, but we don't know how to make that diagnosis of pneumonia, then this is your problem."

6. [Clinician; theme: value of diagnosis quality] -

"For me, as an attending of trainees, I think [diagnosis] is another one of those, just like performance improvement, goals we should all aim to strive for."

7. [Clinician; theme: value to quality improvement efforts] -

"This is critical, I may say, for anything that we do on quality... does the patient really have pneumonia or not."

8. [Clinician & clinical practice guideline author; theme: value to antibiotic stewardship] -

"If your main goal is to reduce the use of antibiotics...if you say, "OK, we are calling almost everything pneumonia" can we really reduce the inappropriate use of antibiotics?"

9. [Clinician & pneumonia quality leader; value to standard pneumonia care] -

"Every ER has access, even walk in. Clinics can do chest X rays. It's not access I I think it's- If you see variability, it's gonna be how some strange practices or different practices getting engrained."

3.1 Contributions Towards Closing Care Gaps

Pneumonia carries a disproportionate burden on older patients as well as patients from under-represented minority groups and with lower socioeconomic status.⁷⁶⁻⁷⁸ Prior studies have also established differences in pneumonia process and outcome measures.⁷⁹⁻⁸² The proposed eCQM would therefore contribute to efforts to advance health equity by supporting quality measurement and improvement of pneumonia processes and outcomes to address these previously known disparities.

We also explored differences in the eCQM's measure performance across different patient groups within the VA and UU healthcare systems. Average measure score on the proposed eCQM (number of positive chest images divided by count of patients with a discharge diagnosis of pneumonia and treated with antimicrobials) and standard deviation were calculated by age, race, sex, hospital complexity level, and rurality of hospital within the full 2015-2022 VA dataset of 89,767 patients with diagnosis and treatment for pneumonia. For the VA system, performance on the proposed eCQM was slightly lower for rural (89.4% with chest imaging) vs. non-rural (90.3%) patients, the oldest age group (>80 years), and women (Table 21, page 16 of *Supplemental Materials* attachment). However, a slightly higher proportion of patients had chest imaging confirmation among non-white patients, while the proportion of patients with confirmatory imaging was similar across complexity of hospital where care was sought. We saw similar patterns in the UU system (Table 21: slightly lower performance among patients >80 years and women; higher performance in non-white patients). All differences, however were <1%.

In summary, while the reliability level of individual facilities appears to vary based on rurality and hospital complexity, at the individual patient level, we did not find large disparities in the measure based on age, sex, race, or rurality of residence. It is important to note that while we had access to a large patient population from the VA and UU systems and we found similar results at these two settings, the populations and healthcare settings studied are not representative of the entire United States, and our findings may be very different for patients in different healthcare systems.

4.1 Feasibility Assessment

We implemented and tested the proposed eCQM within 3 health systems to assess: 1) whether all required data elements were routinely generated during the care of pneumonia hospitalizations and 2) which barriers or challenges exist when implementing and extracting the measure from current healthcare data to inform our feasibility scorecard.

Prior to testing, we conceptualized the people, tools, tasks, and technologies necessary to implement the measure using the Systems Engineering Initiative for Patient Safety (SEIPS)⁶⁶ model (see page 4 of *Supplemental Materials* attachment, Figure 4). We contrast the process of measure implementation by traditional chart review (Figure 4, panel 1) with the proposed eCQM (Figure 4, panel 2).

The majority of inpatients with pneumonia present through the emergency department (ED) or urgent care (86%, internal data), or less commonly outpatient clinics (8%), or other inpatient settings (<2%) where they receive an initial diagnosis and treatment. The initial diagnosis can be uncertain, and timely treatment can appropriately take priority over accurate diagnosis. Our prior work demonstrated that 35% of all patients initially diagnosed with pneumonia from the ED have a change in diagnosis, and 20-25% lack initial chest imaging that confirms pneumonia. Once a patient is hospitalized, it is the task of the hospital care team to either verify a presumptive

diagnosis with additional chest imaging or refine the diagnosis if it is not verified prior to discharge. This entire process is represented by the diagnostic cycle (Figure 4, panels 1 and 2, circles). Factors that influence diagnostic performance may include fragmentation of the healthcare systems (i.e., lack of ability to review outside images or diagnoses), competing priorities by inpatient hospital care teams (attention to treatment and outcomes over diagnostic accuracy), availability of imaging technology and specialists to interpret the results, overall organizational safety culture and clinical inertia, and ability to track a patient's clinical response, which can be influenced by EHR usability and fragmentation. These factors can also impact documentation quality, accurate capture of diagnosis and treatment by coding specialists and eCQM analysts later in time.

Data must be generated from the pneumonia care process to create the measure for both traditional chart review and outlined eCQM (Figure 4). **Inpatient providers** must document within the EHR a discharge diagnosis of pneumonia, order antimicrobials that are recorded in the EHR, and obtain chest imaging that generates a result in the form of a report or structured data that is stored in the **EHR**.

All of the data elements are already are part of normal healthcare system processes for care delivery and billing, so documentation of these items for the proposed eCQM do not add any additional burdens for a healthcare system. However, there is significant work on the part of EHR designers, data managers from the accountable entity, and measurement abstraction tools. The main goal of our feasibility assessment was thus to evaluate the feasibility of abstraction data from the EHR, employ measure abstraction tools to the data, and synthesize the data into the measure for reporting.

To successfully conduct **manual chart review** (panel 1), **data managers** have previously been required to identify diagnoses of pneumonia from administrative data. Manual chart abstraction then is conducted by **expert reviewers** who interpret and apply **consensus definitions**. Design of chart review consensus guides require substantial design and training of reviewers. Once established, however, our experience has been that each chart requires approximately 15 minutes to review.

To successfully conduct the eCQM (panel 2), **data managers** need to extract diagnoses of pneumonia, antimicrobial data, and chest imaging result data in the form of structured data or reports. **Measure abstraction tools** are required to facilitate the conversion of data into coded elements based upon **consensus definitions** and synthesize the elements into the measure. The development and installation of these tools requires significant design, mapping, and some adaptation for different EHR's. Once installed, our experience has been that the measure can be employed in under 1 second per chart.

Regardless of the Measure reporting teams are then needed to submit the measure to the public, providers, and patients.

Feasibility Considerations for the Measure Specification

One of our main goals of the measure was to leverage the most reliable, accurate and standardized data elements and to provide a simple, transparent measure that could be consistently extracted, calculated, and interpreted without substantial data resources or expertise. An existing chart-review based measure of inappropriate diagnosis of pneumonia was recently endorsed and included additional criteria including physical signs and symptoms to improve specificity of the diagnosis. We determined early on that the signs and symptoms were often missing from standardized formats in the EHR, and would not be feasible to collect for the majority of EHR systems without major disruption to existing workflows. Since nearly 75% of the cases with an inappropriate diagnosis of pneumonia lacked a chest image using the recently endorsed measure,⁶⁸ we focused on this component of diagnostic accuracy in pneumonia as the most feasible for an eCQM.

Because the majority of existing quality measures for pneumonia focus on CAP, we explored prior and existing measure approaches that require diagnostic coding position or timing of antimicrobials and chest imaging to obtain a measure more specific to CAP. In the VA and UU systems, we tested 12 measure specifications, representing all possible combinations of the following data elements:

Diagnosis of pneumonia:

- 1) at any position
- 2) in the principal/primary position, or principal diagnosis code for sepsis/respiratory failure and secondary code for pneumonia

Antimicrobial treatment:

- 1) at any time during hospitalization;
- 2) within the first 3 days of admission; or
- 3) within the first 3 days of admission

Chest Imaging:

- 1) within 2 days prior or at any time during hospitalization;
- 2) within the 2 days prior and the 2 days after admission

It was feasible to implement all 12 possible definitions in the VA and UU systems (Full results are available in the *Supplemental Materials* attachment, pages 5-6). Among patients with the most inclusive denominator (positive chest imaging at any point, among patients diagnosed at any position, with antimicrobials at any time) over 90% were also in the most restrictive definitions (principal diagnosis code, antimicrobials within 2 days, chest imaging within 2 days- 92% VA and 96% for UU). Measure performance was also very similar using the inclusive versus restrictive definitions (VA - 90% versus 88% 90%; UU - 92% versus 91%). We thus proceeded with the most inclusive measure definition, which required the least amount of interpretation and data with time stamps.

We also explored adding admission source as an additional exclusion criterion. We considered exclusion of patients admitted from outside hospitals since a higher proportion of them lacked available chest imaging data and thus may demonstrate lower performance due to missing data. However, we noted that for the <5% of pneumonia patients this represents, the measure could encourage data integration between hospitals to verify pneumonia diagnoses that may have been verified at a transferring hospital but are listed at discharge from the receiving hospital. In instances where chest imaging only occurs at an outside setting, the measure would encourage the inclusion of these tests into the EHR to justify claiming the diagnosis as one that was treated at the receiving hospital.

Feasibility: Clinical Workflow

Data elements used in the proposed eCQM specification are already routinely collected as part of the care and billing process, so the eCQM should not impact patient-provider interactions, and there should be no additional cost or burden of data collection for the clinical field when generating this eCQM. The clinical field is already tasked with ensuring that imaging from outside facilities is reviewed and documented (a typical expectation of transfers). Extraction for the data elements used for measure performance calculation and reporting does not require individual patient identifiers and no contact with patients for data collection is necessary, so potential threats to patient confidentiality are limited.

Potential unintended consequences of the eCQM on clinical workflows include:

1. Increased pressure to obtain CT scans during a patient hospitalization, thereby increasing the burden and cost of a hospitalization. However, this may result in more accurate diagnosis and reduce readmissions (see construct validity testing).
2. Increase in the use of lung ultrasound (US) in some settings, as this imaging method could

- provide a low-cost or low-burden alternative to CT scans. To be able to use lung US effectively for this measure, the accountable entity would need to ensure standardized reporting and storage of lung US results into the EHR.
3. A greater burden on the clinical field to upload and store outside images in searchable formats, although this may be an important desired consequence on EHR interoperability and health information exchange.
 4. Increased pressure on radiologists to store radiology documentation in searchable or structured formats or use technology to improve extractability of radiologic confirmation. The current measure and optional NLP package are based on existing searchable radiology documentation and do not disrupt clinical workflow.

Feasibility: Measurement Workflow

The greatest source of increased burden would be to the analysts and programmers tasked with implementing eCQM reporting who will need to map EHR data required for the measure to the eCQM concepts according to individual organizations' usual workflows. It was feasible to identify structured data elements that mapped to all concepts needed for the proposed eCQM in standard terminology. However, while availability and accuracy of the diagnosis code and medication administration were excellent and we found that structured data fields for chest imaging results are currently available within the MAT, the UU, and the VA system, they are not consistently used in current practice at the VA or the UU, and accuracy was low due to missing data: Approximately 65% of all chest images positive for pneumonia contained codes for abnormal imaging. We found no structured data for chest imaging results in the University of Michigan system.

Explanation of Feasibility Scorecard

At all 3 sites, we found that the data elements specified by value sets were captured in a structured format within the EHR ("availability"), provided via an authoritative source ("accuracy"), since the codes were assigned based on standard clinical and billing workflow ("workflow" criterion). Chest imaging results were stored in structured data, meeting the "structured format" and "information is from an authoritative source" criteria. However, existing codes were not consistently mapped to the chest imaging report (when present, the codes were accurate but the codes were missing approximately 60% of the time at UU and VA (no query conducted at Michigan)). As a result, the team developed, implemented, and validated an open source, rule-based NLP algorithm in all three EHR instances to support the measure using existing clinical data. Details of NLP feasibility assessment are described below, along with validation results in later responses. The NLP tested in these systems is only one solution; other NLP of chest imaging exists and are available, as are image processing tools that can be applied directly to images in some systems.⁶⁷

We also found that the data elements were all in nationally accepted standard terminology (e.g., ICD-10 codes, RxNorm), with the exception of medication ordering and administration within the

Epic instance with the University of Utah system, where medication data are stored in structured format within the UU EHR but the standard RxNorm coding can omit combination medications. For this setting, our team identified additional antimicrobial orders and administration to ensure accuracy. Epic does provide an RxNorm national terminology mapping. Similarly, the VA does not use RxNorm system-wide, but there is an existing RxNorm mapping system for medications (“standards” criterion of “coded in a nationally accepted terminology *or can be mapped to that terminology standard*”). At the University of Michigan, the medication data elements were not mapped but text-searchable and identifiable from structured data.

Feasibility: Natural Language Processing (NLP) of Imaging Reports

Current clinical workflow has not yet mapped chest imaging to structured concepts within the three systems where the proposed measure was tested. However, chest imaging is universally stored within the EHR, in the form of direct images or within the free-text narratives of radiology reports. We found that chest imaging reports were stored in a searchable, indexed format for all three systems. Reports were accessible to classify using natural language processing (NLP). We thus assessed the feasibility of a supportive NLP algorithm trained to identify pneumonia on chest X-rays and chest CTs. Computing resources and human time feasibility assessment details are available in Tables 7a and 7b in the *Supplemental Materials* attachment (page 7).

NLP systems are often designed for particular datasets in individual institutions, leading to limited reusability and generalizability to other institutions. We thus developed and tested NLP for interoperability across multiple different systems. Building upon an existing ontology for pneumonia³ (i.e., search terms for pneumonia such as of “infiltrate”, “consolidation” “opacity”), a rule-based algorithm to classify documents as positive, possible, or negative for pneumonia, was first developed and tested within the VA healthcare system. It was then implemented as an “off the shelf” product with no customization, tested, customized, and tested again in University of Utah data.⁶⁹ A rule-based approach was selected because of its feasibility, interoperability between settings, and transparency of the rules in the system to allow for scrutiny and customization, as well as minimal annotation required for re-training. While the thresholds used to classify a chest image report as “positive” were established by clinical input, existing literature, and evaluation of performance, the NLP threshold can be further adapted based upon testing and feedback from the field. The NLP tested in this proposal is available for public use using the Python platform (<https://github.com/medspacy/medspacy>). The NLP system implemented at the VA also included NLP of initial and discharge diagnoses and was initially comprised of 782 rules. The customization for the UU system resulted in 34 additional rules.

To assess NLP feasibility from a computing resources and analyst perspective, an analyst at a third health system, University of Michigan, who had not previously extracted radiology reports, used NLP, or had extensive knowledge of the Python language, extracted radiology reports and

implemented the NLP “off the shelf” with no customization.

NLP implementation at the new site had two primary challenges surrounding installation. First, the Michigan analyst needed to install the NLP system on his machine. We had to update the code for one component of the NLP pipeline and update the required version for the Python platform (medspaCy). This enabled us to install and run the software on a local machine. We do not expect these issues to recur since addressing them, but additional changes may need addressing in the future. Second, the installation process of the NLP system was originally tested on a Mac system. Additional complications arose during installation for a Windows machine, due either to dependencies requiring additional work to install on Windows (such as version of C++ installation) or administrator permission issues on the Michigan analyst’s laptop. After four debugging sessions, a zipped Anaconda python environment was shared that had been built on a University of Utah Windows server and worked immediately. After this, running the NLP package went smoothly.

Once the installation challenges were surmounted, running and executing the NLP was straightforward. We provided an example Jupyter notebook showing how the NLP processes and outputs data to the Michigan analyst. We did not need to provide Michigan any further support to run the NLP once the package was installed. Because validity testing at University Michigan yielded a lower sensitivity than University of Utah and VA, we also conducted an error analysis (see *Validity testing*) and have estimated the human time required for customization of the NLP based upon the input from the Michigan system (in progress).

4.2 Attach Feasibility Scorecard

[eQOM-Feasibility-Scorecard_CBE4440e.xlsx](#)

4.3 Feasibility Informed Final Measure

Diagnosis codes for pneumonia:

We began with diagnosis codes for pneumonia found within the VA and UU EHRs and during drafting of the eQOM within MAT added diagnosis codes that were within existing value sets. We published our own value sets as pre-existing value sets were not comprehensive and excluded diagnosis codes identified within our existing EHR data.

We reviewed with our TEP the proportion of patients meeting each of the 12 criteria and the similar measure performances across different diagnostic coding definitions. Several representatives from the TEP cited the infeasibility of using diagnosis code position to identify CAP versus HAP. Thus, we retained the most inclusive definition of the measure possible.

Antimicrobial use:

We initially defined antimicrobials using only antibiotic data. However, after the COVID19 pandemic, a small (<5%) number of patients were diagnosed with pneumonia, had positive viral tests, and received antivirals but not antibiotics. We found it very feasible to identify these medications used to treat respiratory viruses in the EHR and given that we may expect a growing proportion of patients to be identified with viral etiologies of pneumonia in the future, we expanded the antimicrobial list to include antivirals.

Chest Imaging confirmation (Numerator)

Structured data: Data were available within the UU and VA systems but were not sufficiently populated to be sufficient. Thus, for each of the three systems, NLP was required to calculate the chest imaging data element accurately.

NLP: The process of feasibility testing of the NLP generated 4 key refinements, which are now included in the updated resources (<https://github.com/medspacy/medspacy>):

1. We set more specific requirements for installing the pneumonia NLP package. We created a requirements.txt file that contains all of the Python dependencies and package versions to be installed by pip could avoid issues caused by changing dependencies.
2. Test the installation process in multiple environments prior to implementation.
3. Some Python experience is necessary to install the NLP and debug issues such as incompatible dependencies. Alternatively, distributing complete Python environments like we did could be a solution, rather than expecting implementing teams to build it from scratch.
4. While installation was more complicated and time-consuming than expected, it seemed from our perspective that running and executing the NLP was straightforward and only required a simple example.

Additional work is currently in progress to adapt the NLP for the Michigan system.

4.4 Proprietary Information

Not a proprietary measure and no proprietary components

5.1.1 Data Used for Testing

We tested the measure using data from three EHR systems:

1. Department of Veterans Affairs (VA) health system (103 facilities): acute care hospitalizations admitted from ED 1/1/2015 and 3/31/2022. This dataset was used to evaluate encounter-level reliability and validity, accountable entity-level reliability, chart

review validity, and associations with outcomes).

2. University of Utah (UU - 1 facility): acute care hospitalizations admitted from ED 1/1/2015 and 3/31/2022. This dataset was used to evaluate encounter-level reliability and validity.
3. University of Michigan (UM) - 1 facility): all hospitalizations discharged between 9/29/2015 and 12/11/2021. This dataset was used to evaluate encounter-level validity.

For initial validity testing of each of the elements required for the measure (discharge diagnosis code, antimicrobials, and chest imaging reports), we used enriched samples of patients from the VA and UU hospitals with diagnoses of pneumonia identified by either natural language processing⁶⁹ of the clinical notes or diagnosis codes. This allowed us to estimate the sensitivity of diagnosis codes to identify cases of pneumonia.

For final validity testing of the numerator (chest imaging confirmation), we identified weighted random samples of patients meeting criteria for the denominator (discharge diagnosis plus antimicrobials) from VA and UU, and a subset of UM patients who previously underwent validity testing of the previously endorsed chart review-based diagnosis quality measure (CBE #3671).⁵⁶

The final process for cohort identification differed slightly by site due to nuances of how the datasets were generated. A site-specific diagram of study exclusions to arrive at the initial population specified in the measure are provided in Figures 5a-c in the *Supplemental Materials* attachment (pages 8-10). Slight differences between datasets used for different testing components are provided in a table in *Acceptability* item 4.

5.1.2 Differences in Data

Unless otherwise noted, each component of the proposed eCQM was tested separately for reliability and validity in each of the three healthcare systems (Table 8, *Supplemental Materials* attachment pages 11-12).

To test the **encounter-level reliability and validity of each of the data elements**, we conducted a 2-reviewer chart abstraction of a random sample of 100 hospitalizations from the VA and UU hospitals with at least one diagnosis of pneumonia (either initial admission or final discharge diagnosis), as identified by either ICD-10 code or NLP diagnosis algorithm.

To test the **encounter-level validity of the numerator** (chest imaging-confirmed diagnosis of pneumonia and antimicrobial treatment) among a sample of patients meeting criteria for the denominator, we conducted a single-reviewer chart review abstraction among a random sample of 104 hospitalizations from VA and UU (52 per site), and a larger convenience sample of n=831 of adult patients hospitalized with pneumonia and treated with antimicrobials, which were previously

reviewed for validity testing of a chart review measure (CBE# 3671).⁵⁶ For the VA and UU, we reviewed a weighted random sample of 50 hospitalizations from the denominator with positive chest imaging and 52 patients with no positive chest image to ensure reliable estimates of validity.

5.1.3 Characteristics of Measured Entities

The VA health system is an integrated health system that uses CPRS, with approximately 485,500 acute hospitalizations per year across 103 facilities. The University of Utah is an academic medical center that uses an Epic EHR, with approximately 14,500 hospitalizations per year. The University of Michigan, an academic medical center that uses a different instance of the Epic EHR with approximately 49,730 hospitalizations per year. All three sites have interoperable EHRs where data can be transferred across outpatient and inpatient settings within the healthcare system. A description of the healthcare systems is provided in the Supplemental Materials attachment (Table 9, page 12).

5.1.4 Characteristics of Units of the Eligible Population

Random sampling was used for patient encounter reliability and validity testing of data elements, as described in item 4.2.2. Accountable entity level testing used the most recent complete calendar year of data (1/1/2021-12/31/2021) within the VA healthcare system. Measure-level performance testing was completed without sampling, using the full denominator populations at each healthcare system (89,767 within the VA; 3,030 at the University of Utah; 831 at the University of Michigan), at the patient encounter level of analysis. Descriptive statistics of patients included in the patient encounter level testing datasets are provided for the overall cohort and by numerator status (chest imaging consistent with pneumonia present vs. absent) in the tables in the *Supplemental Materials* attachment (VA: Table 10; University of Utah: Table 11; University of Michigan: Table 12). Accountable entity level testing dataset characteristics are provided in the item above (measured entity descriptive characteristics). For the University of Michigan population, abstractors screened consecutive patients via medical record review 30 days after discharge and included the first eligible patient daily, abstracting 8 eligible patients during a two-week cycle. Patients were eligible for inclusion if they were adults (≥ 18) admitted to general care with a billed discharge ICD-10 code of pneumonia, and received antibiotics on day 1 or day 2 of hospitalization. Patients who had documentation of treatment for an additional infection unrelated to pneumonia, were severely immunocompromised, were pregnant, were admitted for comfort measures, or who left against medical advice were ineligible.

5.2.1 Level(s) of Reliability Testing Conducted

Person or encounter level (i.e., data element) (e.g., inter-abstractor reliability), Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)

5.2.2 Method(s) of Reliability Testing

Encounter-Level Reliability Testing (VA and UU)

Among patient hospitalizations at the VA and UU with either an initial or discharge pneumonia

diagnosis identified by ICD-10 code or NLP, a random sample of 100 patient charts was reviewed independently by two board-certified clinician reviewers (emergency department physician pulmonologist) following a previously published standard consensus guide.⁶⁸ Reviewers validated the following elements:

- ICD-10 discharge diagnosis of pneumonia
- Chest image report consistent with pneumonia
- Note: inter-rater reliability testing of RxCUI codes for antimicrobials (antibiotics and antivirals) was not performed as this coding system is the industry standard for pharmacy management, medication administration, drug interaction monitoring software, and associated billing.⁶⁹

Pairwise inter-rater reliability (IRR) between two clinician reviewers was calculated for each healthcare system for each of the data elements using Cohen's kappa.⁷⁰ Additional details on Cohen's kappa is provided in supplemental material, page 15 but, in short it provides a statistical estimate of the extent of agreement between clinician reviewers beyond that expected by chance. There was no missing data and no sensitivity analyses were performed.

Accountable Entity-Level Reliability Testing

"Signal-to-Noise Analysis"

Accountable entity-level reliability testing was assessed by two methods, initially using the Adams protocol⁷¹ detailed below and then using an empirical Bayes approach to account for small numbers of pneumonia cases at some hospitals. Using the 103 acute care facilities and 89,767 hospitalizations diagnosed and treated for pneumonia present in the Site 2 (VA) population, the 2015-2022 full cohort was restricted to a single year of the most recent data (2021) to assess reliability using a one-year reporting period as specified for this proposed eCQM. This most recent year of data included 100 VA hospitals and 8,253 hospitalizations with a discharge diagnosis of pneumonia where the patients also received antimicrobials (measure denominator). The same dataset restricted to a single year of pneumonia diagnosis and treatment in the VA was used for both approaches used to estimate measure reliability.

Approach 1: Adams Protocol

Reliability testing at the accountable entity level was initially performed according to the Adams protocol.⁷¹ A facility level beta binomial regression model was used to estimate intermediate parameters, alpha and beta, using the number of hospitalizations meeting the eCQM denominator definition (i.e. hospitalizations with pneumonia diagnosis treated with antimicrobials, "n trials") and the eCQM numerator (the subset number of hospitalizations in the eCQM denominator with chest imaging consistent with a pneumonia diagnosis) for each facility. The output of this regression model, parameter estimates for alpha and beta, were then used in a formula to

calculate inter-facility variance, where interfacility variance is: $(\alpha * \beta) / [(\alpha + \beta + 1) * (\alpha + \beta)^2]$.

This interfacility variance represents the amount of “true” variation in measure score performance between the VA hospitals. Once the single year inter-facility variance was estimated for the cohort, it was used to calculate facility-specific estimates of reliability using the formula²:

$$\text{Reliability} = \frac{\sigma^2_{\text{inter-facility}}}{\sigma^2_{\text{inter-facility}} + [p(1 - p) / n]}$$

where p = facility-specific eCQM numerator / facility-specific eCQM denominator and, n is the facility-specific eCQM denominator. The $[p(1-p) / n]$ portion of the formula is facility-specific (within facility variance) due to measurement error. Thus, the denominator of the reliability calculation represents the total variance in eCQM score for VA hospitals in 2021.

Facilities were then sorted by size of eCQM denominator (i.e., number of hospitalizations with discharge diagnosis of pneumonia and treatment with antimicrobials) and divided into deciles based on denominator size. The mean reliability score and standard deviation was then calculated for all facilities within a decile category.

This analysis provides a “signal-to-noise” estimate, specifically the proportion of the total measure variance that is due to “true” differences in proportion of diagnostic concordance between hospitals. Reliability metrics provided across each decile of facility size permits assessment of how well the measure differentiates between different hospitals’ performance on obtaining chest imaging consistent with pneumonia for all hospitalized cases diagnosed and treated for pneumonia, across the range of facility sizes.

Approach 2: Empirical Bayes

To assess whether reliability estimates produced through the Adams protocol were potentially impacted by unstable mean estimates from facilities with small denominator sizes and similar measure scores between first and tenth deciles of performance, an empirical Bayes method⁷² was used. In this approach, information on the overall mean distribution of the measure performance score data is used to inform estimated scores on the eCQM for facilities with smaller denominator sizes and is similarly used to calculate weighted variance estimates for the small denominator facilities prior to calculating the “signal to noise” estimate.

Patient and Facility Characteristics Related to Reliability

To examine whether certain patient or facility characteristics were related to reliability level, the person and facility characteristics reported in the measured entity and patient encounter level tables in the *Scientific Acceptability* section above were analyzed for facilities in the first (lowest) decile of reliability and tenth (highest) decile of reliability.

5.2.3 Reliability Testing Results

Encounter-level reliability testing (Table 13), accountable entity reliability testing by two different decile grouping strategies (Tables 14 and 15), and an exploration of patient- and facility-level characteristics associated with high vs. low reliability decile (Tables 16a and 16b) are available in the Reliability Testing Results_CBE4440e attachment in Item 4.2.3a. Table 14 and the manual entry table for Item 4.2.3a table contain the same results.

5.2.3a Attach Additional Reliability Testing Results

[Reliability Testing Results_CBE4440e.pdf](#)

5.2.4 Interpretation of Reliability Results

We found approximately one quarter of hospitalizations with either an initial or discharge diagnosis of pneumonia and treatment with antimicrobials for pneumonia had a hospital discharge diagnosis of pneumonia captured by ICD-10 code at any position (Table 13). Inter-rater reliability for the presence of a pneumonia ICD-10 code (any position) was excellent across health systems, ranging between 0.88 and 0.91.⁷⁰ This indicates that potentially only a fraction of pneumonia diagnosis is captured by coded data but among hospitalizations with coded encounters, there is high reliability of the pneumonia diagnosis code. Inter-rater reliability was also high for assessing presence of chest imaging and that imaging was consistency with a diagnosis of pneumonia (0.87 both healthcare systems). Because receipt of antimicrobials is known to be highly accurate,⁷³ we did not assess inter-rater reliability.

The diagnosis of pneumonia can be subjective, and interpretation of both chest images and their reports can demonstrate uncertainty and low reliability, even between two radiologists.⁷⁴ However, our assessment suggests that when following a consensus guide with established definitions,⁶⁸ two reviewers can reliably (1) identify bedside clinician diagnoses of pneumonia and (2) classify chest images. Radiology results are higher than other evaluations of the reliability to identify conditions on chest imaging reports.⁷⁵ We would not expect similar results among clinicians who are not provided training on established definitions.

As shown in Table 14, a smoothed mean reliability that accounts for unreliable mean estimates from hospitals with small numbers of annual pneumonia diagnoses treated with antimicrobials

results in higher estimates of mean reliability by facility size deciles. Mean reliability across all facilities was 38% by the Adams protocol and 62.2% when accounting for small facilities with empirical Bayes, which is considered “moderate reliability”. Reliability estimates are impacted by the similar performance scores across facilities which reduces variation in score due to “signal” compared to overall variance with a binary outcome. Table 15 provides reliability estimates for both estimation methods for deciles grouped by facility reliability score rather than denominator size. Here the reliability ranges between 35% and 100%, with half of facilities meeting or exceeding the moderate reliability threshold with the empirical Bayes approach. Uncorrected for small denominators, the reliability ranges from 1.9% to 100% using the Adams approach. The proposed eCQM as specified appears to have moderate reliability. It should also be noted that this analysis was conducted among facilities all within the same healthcare system and thus may be more similar in performance to one another than a true population distribution of all hospital types and systems within the United States.

Patient and Facility Characteristics Related to Reliability

As shown in Tables 16a and 16b, the patient and facility characteristics associated with low and high reliability were explored. The primary characteristics associated with reliability level were rurality of patient residence or geographic hospital location and complexity of the facility where patient was hospitalized. There were some treatment differences between high and low reliability facilities as well, including higher probability of being treated with an antimicrobial or diuretic within the first 24 hours of hospitalization.

Table 2. Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size

Accountable Entity-Level Reliability Testing Results													
	Overall	Minimum	Decile_1	Decile_2	Decile_3	Decile_4	Decile_5	Decile_6	Decile_7	Decile_8	Decile_9	Decile_10	Maximum
Reliability	62.2	35.0	59.9	54.0	54.1	58.7	63.6	62.4	66.1	64.4	68.9	69.8	68.1
Mean Performance Score	92.4	83.3	92.0	91.8	89.7	93.1	92.6	92.8	94.0	92.2	94.1	92.0	87.6
N of Entities	100	1	10	10	10	10	10	10	10	10	10	10	1
N of Persons / Encounters / Episodes	8253	6	236	356	467	564	657	748	893	1098	1290	1944	283

5.3.1 Level(s) of Validity Testing Conducted

Person or encounter level (i.e., data element) (e.g., sensitivity and specificity), Accountable entity level (i.e., measure score) (e.g., criterion validity)

5.3.3 Method(s) of Validity Testing

Encounter-Level Validity Testing

Criterion Validity of eCQM Components

- ICD-10 discharge diagnosis of pneumonia (all 3 health systems)
- Chest image report consistent with pneumonia
- RxCUI codes for antimicrobials (antibiotics and antivirals)

For the VA and University of Utah, we conducted two chart review validations:

1. Individual data elements of diagnosis code and chest imaging in a sample of 100 cases (50 from each system) by 2 board-certified clinician reviewers each with an initial or discharge diagnosis of pneumonia (using ICD codes and NLP) to evaluate validity of the ICD codes among an enriched sample of patients with pneumonia diagnoses documented in the chart. Performance characteristics (sensitivity, positive predictive value, and negative predictive value) were calculated for each data element.
2. The full measure was validated in samples of patients meeting denominator criteria by a single clinician reviewer with a total of 104 cases at the VA (52) and UU (52), with a weighted random sample of 25 hospitalizations with chest imaging results consistent with pneumonia and 26 hospitalizations without chest imaging consistent with pneumonia equal sample of numerator positive (25) and numerator negative (26) cases. Given that the ratio of chest imaging-positive to negative cases is 9:1 (performance of 90%), we then calculated performance characteristics using weighted sampling.

For both validations, clinician reviewers were trained then followed a published consensus guide with established definitions for positive classification threshold. Chart review guide is publicly available in the literature⁶⁸ and on GitHub (https://github.com/abchapman93/medspacy_pneumonia).

For the University of Michigan, 831 cases previously reviewed by chart abstractors who had a discharge diagnosis of pneumonia and received antimicrobials were used. As this second validation was performed within a subpopulation restricted to one that had a discharge diagnosis of pneumonia and received antimicrobials, only positive predictive value (not sensitivity) was calculated for this second criterion validation. Chart review guide for UM chart-review measure is available publicly (<https://mi-hms.org/inappropriate-diagnosis-community-acquired-pneumonia...>).

Document-level Criterion Validation of Chest Image Report Data (NLP)

Original NLP development and validation is previously reported.⁶⁸ Briefly, two board-certified clinical reviewers (an emergency department physician and a pulmonary and critical care physician) annotated notes and developed consensus guidelines by iteratively reviewing new groups of 10-20 imaging reports, tracking inter-rater reliability, and coming to consensus on disagreements. After reaching an inter-rater reliability of 0.8, the physician annotators began to single rather than double annotate the sampled documents for the testing set. Chest imaging NLP performance was assessed using sensitivity (recall), positive predictive value (precision) and F1 ((PPV*Se) / (PPV + Se)).

For the VA document-level validation, 300 notes were double-annotated with disagreements

reviewed and resolved by consensus. The final IAA on the testing set was 94.9. For the UU dataset, clinicians annotated a total of 267 documents for development, with 91 double-annotated to measure IAA of 0.89. One clinician then annotated 100 additional documents from each setting for testing. UU testing notes were not double-annotated, as the high IAA established confidence in annotator reliability using the UU dataset.

Accountable Entity Level Criterion Validity Testing

Overall measure performance

- eCQM numerator measure of concordance between chest imaging result and pneumonia diagnosis + treatment with antimicrobials (chest imaging consistent with pneumonia AND discharge diagnosis of pneumonia along with receipt of antimicrobials during hospitalization)

Measure score validity was assessed within all three healthcare systems by comparing the EHR-derived determination of whether chest imaging-consistent with pneumonia was present with chart review determination of its presence as the gold standard. Since this was a validation of the score performance characteristics, the validation was conducted within a dataset restricted to patients who met eligibility criteria for the denominator (discharge diagnosis of pneumonia and received antimicrobials). Thus, score validation was conducted among hospitalizations with a discharge diagnosis of pneumonia as identified by ICD code alone and receipt of eligible antimicrobials. A random sample of 25 concordant (chest imaging consistent with pneumonia present by measure specification) and 26 discordant (no chest imaging consistent with pneumonia by measure specification) in the VA dataset were manually chart reviewed to determine whether measure specified designation of meeting/not meeting criteria were correct. This was repeated with the same number of charts in the corresponding University of Utah denominator dataset. Discordant cases occur within approximately one in ten charts. To assess enough eCQM-identified discordant hospitalizations for validation, a weighted sample of equal number of measure discordant hospitalizations (no evidence of chest imaging consistent with pneumonia) relative to concordant was obtained. The raw and then correction for weighted sampling results are provided. Validation was also conducted within the University of Michigan dataset of 831 patients with antimicrobial treatment and discharge diagnosis, along with full chart review.

Association with Outcomes: Construct Validity

- Association of eCQM score and key processes and outcomes for pneumonia.

At the accountable entity level, we evaluated crude relationships between performance score and key process and outcome measures among 100 VA facilities with >250 pneumonia hospitalizations. We calculated the Pearson correlation between the process or outcome measure

vs. performance score and visualized facility-level relationships between the performance score and measures using scatter plots.

At the individual level, we assessed patient-level associations between measure performance and with the same key measures. A logistic regression model was fitted for each of the outcomes for the discordant group and for its comparison- patients with a negative chest image who did not receive a diagnosis of pneumonia – based upon patient characteristics. The covariates of the logistic models included facility ID, discordance status and 27 patient variables that included demographics comorbidities, vital signs and labs. For each patient in the discordant group, we predicted the counterfactual probability of each outcome were his/her discordant status to change to the comparison status but with the other covariates remaining constant. The mean of the predicted probability was calculated and referred to here as the counterfactual marginal mean. The observed marginal mean was calculated as the observed probability of the outcome in the same group of subjects. For each patient, we calculated the difference between the observed versus the predicted counterfactual probability of each outcome; that is, the probability of the outcome that would be predicted were the patient to have diagnostic concordance. The observed marginal means and predicted marginal means were then compared.

5.3.4 Validity Testing Results

Please see tables 17 through 20 and figures 6 and 7 in the 4_3_4_Additonal Validity Testing Results_CBE4440e attachment. Accountable entity-level scatterplots for all key processes and outcomes assessed are shown in the 4_3_4_Additonal Validity Testing Results_CBE4440e attachment (Figure 8, pages 6-7).

5.3.4a Attach Additonal Validity Testing Results

[4_3_4_Additonal Validity Testing Results_CBE4440e.pdf](#)

5.3.5 Interpretation of Validity Results

Criterion Validity: Diagnoses

As illustrated in Table 17 (4_3_4_Additonal Validity Testing Results_CBE4440e attachment, page 1), discharge diagnosis codes were not very sensitive, identifying approximately half of the total pneumonia cases identified at the UU and VA systems through a combination of ICD codes and NLP of clinical notes. Diagnosis codes had excellent positive predictive value for VA (100%) but only fair PPV for UU (75%). With the addition of antimicrobials, however, the PPV of diagnosis codes and antimicrobial increased to 100% for both systems.

Criterion Validity: Antimicrobials

The presence of an electronically obtained measure of antimicrobial use demonstrated a PPV of 100% among all the samples tested (Table 18a, 4_3_4_Additonal Validity Testing

Results_CBE4440e attachment). Prior work has established a very high sensitivity among all systems and thus was not repeated.

Criterion Validity: Chest imaging

At the document level among chest images reports in hospitalized patients, the NLP algorithm was highly sensitive (near 100%) at finding chest imaging consistent with pneumonia and had moderate to high PPV, depending upon the population (68-100%). PPV increased to 100% when applied to patients meeting criteria for the measure denominator (Final measure validation, Table 18). At the patient hospitalization level, however, sensitivity was slightly lower. Performance of the “off the shelf” and customized NLP approaches are shown in Table 19 (4_3_4_Additional Validity Testing Results_CBE4440e attachment).

Criterion Validity: Complete measure

Overall measure score validity (Table 18a) was highly sensitive for finding concordant patients (chest imaging consistent with pneumonia) among patients diagnosed and treated for pneumonia at VA (98.2%) and UU (97.8%). However, sensitivity was slightly lower among the previous chart-reviewed sample of hospitalizations from the University of Michigan (90%). An error analysis at the University of Michigan is underway; preliminary results have revealed additional key terms used for pneumonia that were more common in their reporting (i.e., “ground-glass”) as well as a slightly different threshold for abnormal classification defined by the reviewer team (for example: “opacities consistent with pulmonary edema” were accepted as positive by reviewers but negative by NLP). Customization of NLP is in progress. PPV of the final measure was near-perfect across three different health systems (100% in VA and University of Utah; 98% at University of Michigan).

Overall, the validity testing suggests the NLP algorithm for classifying chest imaging performed well even at the third site which had no prior NLP experience, no customization, and a different team of clinician reviewers with slightly different consensus definitions for positive image results. Performance can be further improved with universal establishment or dissemination of consensus criteria.

Construct Validity: Patient and Accountable Entity Association with Outcomes

At the **individual patient level**, we did not observe any clinically significant associations between the measure performance and clinical outcomes (Figures 6 & 7; Table 20; 4_3_4_Additional Validity Testing Results_CBE4440e attachment). However, we did observe a strong association between lower performance and antibiotic use. Among patients with a pneumonia diagnosis and negative chest imaging, the observed antibiotic use rate was 44% greater than the antibiotic use expected for similar patients not receiving a pneumonia diagnosis.

At the **facility level**, VA facilities with higher performance demonstrated lower hospital re-admission rates among all hospitalized patients ($R=-0.22$, 95% CI: [-0.40, -0.02]) but not among pneumonia patients, and a higher proportions of patients receiving a CT among pneumonia hospitalizations ($R=0.31$, 95% CI: [0.16, 0.45]) as well as among all hospitalizations ($R=0.34$, 95% CI: [0.19, 0.50]). We also observed a non-significant correlation between facility-level measure performance and guideline-concordant empiric antibiotics as well as pneumonia mortality, but no correlation between performance and antibiotic use or mortality among all hospitalizations.

Overall, it is difficult to discern whether the observed associations represent causal relationships or merely associations related to patient differences, particularly for individual patient-level associations. However, these analyses may suggest that if performance of this measure improves, we may observe a reduction in unwarranted antibiotic use at the patient level and an increase in CT scans for both pneumonia and non-pneumonia hospitalizations. It is unclear whether the reduction of all-hospital readmissions could be anticipated, but a causal relationship between performance and this outcome could possibly be mediated by more thorough diagnostic workup (such as CT's) and an improvement in diagnostic accuracy.

5.3.2 Type of Accountable Entity Level Validity Testing Conducted (derived)

Empirical validity testing at the accountable entity-level (e.g., criterion validity, construct validity, known groups analysis)

5.4.1 Methods Used to Address Risk Factors

No risk adjustment or stratification

6.1.1 Current Status

Not in use

6.1.2 Current or Planned Use(s)

Public Reporting, Payment Program, Quality Improvement (Internal to the specific organization)

6.2.1 Actions of Measured Entities to Improve Performance

We explored how measured entities would improve measure performance from multiple stakeholder perspectives through three activities: 1) 5 meetings with Technical Expert Panels (TEPs); 2) usability testing with clinicians interacting with measure feedback at the University of Utah; and 3) planning for accountability applications after initial endorsement, including review with representatives from the eQIM team at the University of Utah.

Technical Expert Panel

The technical expert panels (TEPs) were each comprised of two patients with experiences of pneumonia and misdiagnosis, as well as several clinicians who were selected based upon their

experience and expertise as informaticians, public health officers, health system leaders, measure developers, and researchers with expertise in pneumonia and healthcare quality measurement. Full details surrounding TEP methods and focus group format are described in the *Supplemental Materials*. Briefly, meetings were conducted via videoconferencing (four during the initial measure development phase and an additional TEP during final measure specification of proposed eCQM (Table 6, page 3 of *Supplemental Materials* attachment). For all five panels, an agenda with sets of specific questions tailored to the measure development phase were provided to panelists prior to the discussion session. Following a presentation by the eCQM development team, moderators unaffiliated with the measure development process facilitated an hour feedback session about the measure. For the development phase TEPs, post-meeting surveys with additional questions were completed by panelists to elicit additional feedback.

Clinicians from the TEP expressed that a measure surrounding chest imaging-confirmation of pneumonia diagnosis and treatment could be used for internal quality improvement to improve diagnosis through feedback. They also noted that the measure aimed at improving diagnosis at a system level, rather than telling individual providers they are “right” or “wrong” at a given point during an episode of care, would be more likely to be accepted by clinicians.

TEP participants identified two important tasks needed to ensure usability of the measure 1) reducing clinician skepticism of the measure through transparency and consensus definitions, and 2) using more than educational interventions to change behavior. The clinician with expertise in pneumonia quality improvement suggested that improving the consistency of a definition of pneumonia would ensure that current point-of-care pathways for pneumonia are used for the appropriate patients. The clinical informaticist pointed out that the measure would only be effective at improving diagnosis if it was integrated into provider workflow in the form of decision support to “make it easy to do the right thing.” Clinicians expressed that access or resources chest imaging should not be a barrier since availability of imaging is universally available in emergency departments and hospitals. Greater use of CT scans were discussed, which was considered a diagnostically superior approach.⁸³ For inpatients, cost of X-rays versus confirmatory CTs were discussed, but participants felt those costs were outweighed by the benefit among “a very highly selected group of patients that require an admission to the hospital rather than everyone with community-acquired pneumonia.” The TEP discussed that there may be more challenges to implementation in outpatient settings, where cost and patient access to imaging may be a bigger factor for considering unintended consequences. TEP participants felt that eCQM scores could be influenced by some patient factors related to risk of pneumonia (e.g., health systems with a higher proportion of smokers), provider knowledge and attitudes toward diagnostic accuracy and imaging, integration of medical records, and organizational culture.

Patients shared experiences having difficulty communicating about diagnoses with their providers and felt that if the chest imaging concordance eCQM made it easier for doctors to diagnose or treat patients and communicate the confirmation with their patients, they were “in full support”. They reported the role of objective findings including chest imaging results to

patient-provider communication, and they suggested that the proposed eCQM could improve care by empowering patients to communicate better with providers, especially around the amount and sources of uncertainty inherent in a pneumonia diagnosis. One patient recalled that the absence of objective findings or communication of uncertainty in previous encounters damaged his trust in clinicians, and that the measure could facilitate such conversations and thus improve patient confidence in providers' decision-making.

In a post-TEP follow-up survey, one respondent felt that the measure and associated mockup of a type of EHR-embedded report of the measure could improve diagnostic accuracy or patient care in the emergency department "if it is a standard that all providers will use to diagnose and treat." Detailed TEP summaries and illustrative quotes are available in the supplemental attachment. Illustrative quotes from panelists directly related to perceived usability of the proposed eCQM are provided in Table 22 (page 17 of *Supplemental Materials* attachment).

Usability Testing

We explored usability by evaluating clinician reactions to pneumonia diagnostic quality measurement through a feedback tool. During early development and evaluation of the measure, an EHR-based dashboard that reported annual provider- and facility-level performance on the proposed eCQM was developed and piloted with emergency department physicians to evaluate the initial ED diagnosis. Linking performance summaries to individual cases was a key feature that enabled individual providers to 1) verify and thus trust the measure results and 2) immediately review individual cases for continued learning. The tool could be either linked with or incorporated into a facility's EHR to reduce burden for retrospective case review. Results could also be summarized annually at the entity level during division meetings along with other process measures and review of the most current clinical practice guidelines.

Planning for Implementation

To improve performance on the currently proposed imaging-diagnosis concordance measure, accountable entities would need to 1) encourage providers to obtain and document chest imaging consistent with pneumonia that they are treating for suspected pneumonia, or 2) decrease the number of patients who are inappropriately diagnosed with pneumonia but have alternative diagnoses such as bronchitis or COPD exacerbations. Accountable entities would need to implement interventions targeted to inpatient clinicians. Prior quality initiatives for pneumonia care have shown success with both audit and feedback⁸⁴ and real-time decision support.⁸⁵ Both of these initiatives require clinician trust and agreement with the measure as representative of quality. If integrated well into clinical workflow, eCQMs have the unique ability to simultaneously measure and improve quality through real-time decision support ("Measure-ventions").^{86,87}

The currently proposed chest imaging-pneumonia diagnosis measure is carefully framed as a measure of concordance between diagnosis and clinical data, rather than as physician “error,” based on feedback from a TEP participant to encourage buy-in from clinicians. The measure is also reported at the accountable entity level rather than individual providers, compatible with additional TEP feedback that receptiveness to the measure would be increased by framing the measure as a performance improvement tool rather than provider-level punitive monitoring. Furthermore, the measure is carefully defined to follow diagnostic guidelines for pneumonia and improving care processes for patients with pneumonia, while avoiding penalizing clinicians for the complexity and diagnostic uncertainty inherent in pneumonia.

If endorsed, there are several potential systems in which the proposed measure can be implemented and evaluated prior to the 1st maintenance review in 5 years in 2029 to provide information for maintenance evaluation.

The University of Utah Health Operations. The quality reporting team will use the eCQMs when the Epic eCQM dashboard functionality becomes available. Per the University of Utah operations team, Epic develops and releases these dashboards whenever an eCQM is incorporated within the CMS Inpatient Quality Reporting program and has additional in-development measures awaiting final inclusion in the Inpatient Prospective Payment Systems Final Rules. The Epic dashboard is largely based on zipped files provided from the MAT development process and is included with the base inpatient Epic package (“Inpatient Clinical Documentation”). University of Utah has previously embedded an eCQM for patients with stroke into point-of-care decision support. As the currently proposed eCQM does not require any Epic specialty apps to capture documentation that best implements workflows, a relevant eCQM dashboard should be available to a wide range of hospitals using Epic if the measure is adopted by CMS. We will work within the University of Utah to integrate the measure within clinical workflows as decision support, similar to the existing stroke eCQM.

Intermountain Health System. Intermountain Health has successfully implemented and tested several versions of real-time pneumonia decision support which includes screening for diagnosis in the emergency department that previously used NLP to identify pneumonia on chest imaging, with similar thresholds to the NLP tested in this proposal.^{88,89} They currently utilize image processing tools to identify pneumonia directly from chest images, also with definitions to the NLP.⁴ They are in the process of evaluating barriers to scaling their decision support to other systems. (<https://classic.clinicaltrials.gov/ct2/show/NCT06008314>) If the measure is endorsed, we will collaborate with this team to 1) use the measure retrospectively to evaluate the impact of decision support on diagnostic quality, and 2) examine the feasibility of real-time decision support that incorporates the measure prior to hospital discharge.

The University of Michigan and HMS. The proposed measure has been successfully

implemented within this system (see *Validity* testing). This testing revealed slightly different thresholds for positive chest imaging that was used compared to the UU and VA systems, highlighting the importance of establishing a consensus definition that can be shared across systems. We will continue to explore ways to optimize the accuracy of the measure in this system and explore ways to augment or support their existing quality measurement and improvement program.

In summary, we explored the usability of the measure to improve pneumonia diagnosis and management from multiple perspectives. Key themes included the importance of 1) focusing on system-level measurement of diagnostic quality; 2) integrating the eCQM within clinical workflow as a form of decision support in addition to feedback that is transparent and useful to clinicians; 3) encouraging objective findings such as chest imaging to promote communication and trust with patients about diagnosis. A key challenge to usability includes the widespread dissemination of a consensus definition of the measure that can be applied consistently across all systems and trusted by clinicians and accountable entities. Given the importance of clinician trust in this measure as one that represents a shared standard, the measure would be most successful if there is substantial dialogue and input on the definition of chest imaging-confirmed pneumonia. If the measure is endorsed, we will engage with additional stakeholders including professional radiology societies, informatics representatives including NLP and image processing developers, and quality measure representatives to establish and disseminate consensus definitions to ensure the credibility, usability, and transparency of the measure.

7.1 Supplemental Attachment

[2. Supplemental Materials Attachment for PQM_CBE4440e.docx](#)

Developer POC email

lindsay.visnovsky@hsc.utah.edu

Measure Developer POC

Lindsay Visnovsky
University of Utah
295 Chipeta Way
Salt Lake City, UT 84108
United States

The measure developer is different from the measure steward

Yes

Steward Address

Michael Good
University of Utah
201 Presidents Cir

Rm 203
Salt Lake City, UT 84112
United States

Steward Organization

University of Utah

Steward Organization URL

<https://www.utah.edu>

Steward POC email

michael.good@hsc.utah.edu