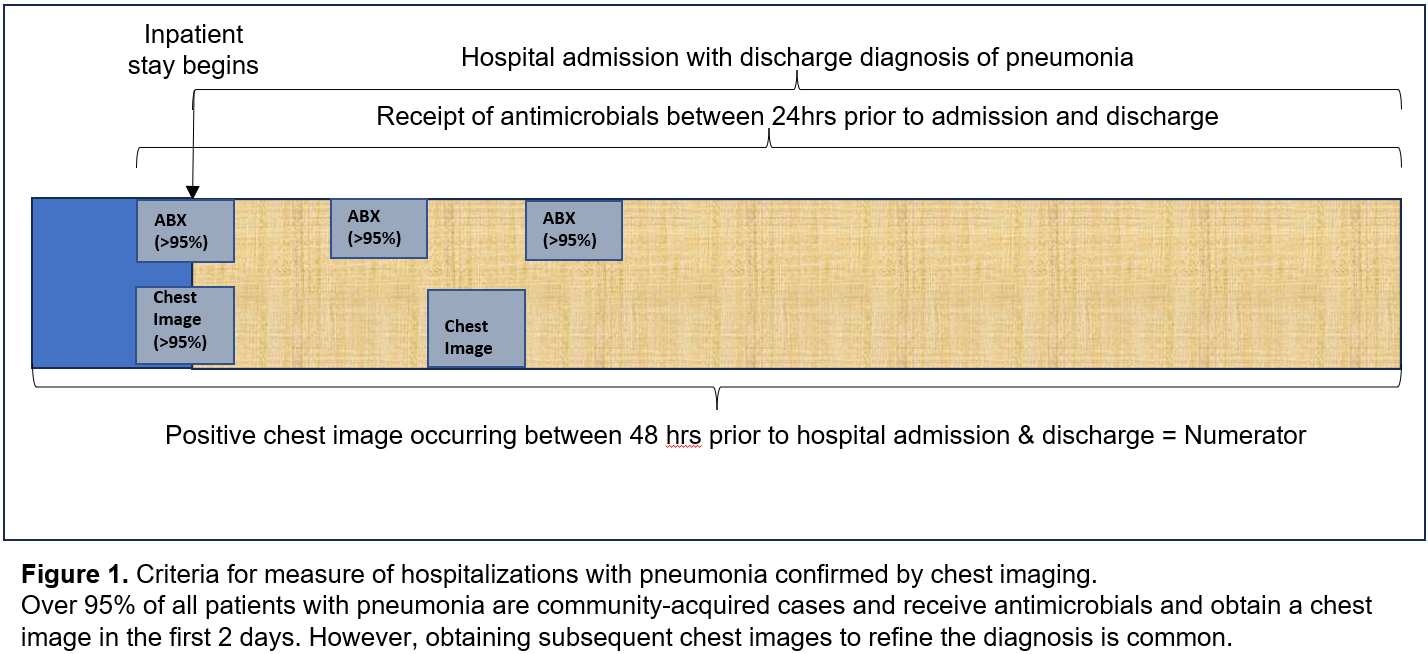
Supplemental Tables for PQM CBE #4440e

**1.14a -- Numerator Details:**



**2.5 – Health Care Quality Landscape:**

Table 5. Sample of Current and Past Pneumonia Clinical Quality Measures.7,45,50,56,60

| **Measure Name** | **Focus** | **Agency** | **Status** |
| --- | --- | --- | --- |
| **NONE** | **Diagnosis** | **Electronic Clinical Quality Measures for CMS Quality Reporting Programs for Eligible Hospitals & Critical Access Hospitals** | **None** |
| 01508-02-C-MIPS – Simple Pneumonia with Hospitalization | Cost | CMS Merit-Based Incentive Payment System (MIPS) | Active |
| PN-2 – Pneumococcal Vaccination | Prevention | The Joint Commission / CMS MIPS Value Pathways | Retired |
| 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 | Diagnostic | The Joint Commission | Active |
| PN-3b – Blood Cultures Performed in the Emergency Department Prior to Initial Antibiotic Received in Hospital | Diagnostic | The Joint Commission | Retired |
| PN-4 – Adult Smoking Cessation Advice/Counseling | Prevention | The Joint Commission | Retired |
| PN-5 – Antibiotic Timing (Median) | Treatment | The Joint Commission | Retired |
| PN-5a – Initial Antibiotic Received Within 8 Hours of Hospital Arrival | Treatment | The Joint Commission | Retired |
| PN-5b – Initial Antibiotic Received Within 4 Hours of Hospital Arrival | Treatment | The Joint Commission | Retired |
| PN-5c – Initial Antibiotic Received Within 6 Hours of Hospital Arrival | Treatment | The Joint Commission | Retired |
| PN-6 – Initial Antibiotic Selection for Community-Acquired Pneumonia (CAP) in Immunocompetent Patients | Treatment | The Joint Commission | Active |
| PN-7 – Influenza Vaccination | Prevention | The Joint Commission |  |
| IQI 20 – Pneumonia Mortality Rate | Mortality | Agency for Health Care Research and Quality Inpatient Quality Indicators (IQI) |  |
| CBE 0147 – Initial antibiotic selection for community-acquired pneumonia (CAP) in immunocompetent patients | Treatment | Centers for Medicare & Medicaid Services | Endorsement removed 2017 |
| CBE 0096 – Community-Acquired Bacterial Pneumonia (CAP): Empiric Antibiotic | Treatment | Physician Consortium for Performance Improvement | Endorsement removed 2017 |
| CBE2882 - Excess days in acute care (EDAC) after hospitalization for pneumonia | Length of Stay | Centers for Medicare & Medicaid Services | Endorsed  2021 |
| CBE 2579 Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia | Outcomes | Centers for Medicare & Medicaid Services | Endorsed 2021 |
| CBE 0506 – Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization | Outcomes | Centers for Medicare & Medicaid Services | Endorsed 2021 |
| CBE 3671 – Inappropriate diagnosis of community-acquired pneumonia (CAP) in hospitalized medical patients | Diagnosis | University of Michigan | Endorsed 2021 |
| CBE 2579 – Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia (PN) | Cost | Centers for Medicare & Medicaid Services | Endorsed 2021 |
| CBE 0279 Community Acquired Pneumonia Admission Rate (PQI 11) | Incidence | Agency for Healthcare Research and Quality | Endorsement removed 2021 |
| CBE 0140 – Ventilator-associated pneumonia for ICU and high-risk nursery (HRN) patients | Incidence | Centers for Disease Control and Prevention | Endorsement removed 2012 |
| **Measures for other conditions that may influence or be influenced by proposed eCQM** | | | |
| CBE 0500 -- Severe Sepsis and Septic Shock: Management Bundle / SEP-1 | Diagnosis/Treatment | Henry Ford Hospital | Endorsed 2021 |
| CBE 0058 – Avoidance of Antibiotic Treatment in Adults With Acute Bronchitis (AAB) | Treatment | National Committee of Quality Assurance | Endorsed 2021 |
| CBE 1891 - Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization | Outcome |  | Endorsed 2021 |
| CBE 1893 – Hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization | Outcome | Centers for Medicare & Medicaid Services | Endorsed 2021 |
| CBE 2859 Pharmacotherapy Management of COPD Exacerbation | Treatment | National Committee of Quality Assurance | Endorsed 2021 |

**2.6 – Meaningfulness to Target Population:**

Technical Expert Panel (TEP) methods details:

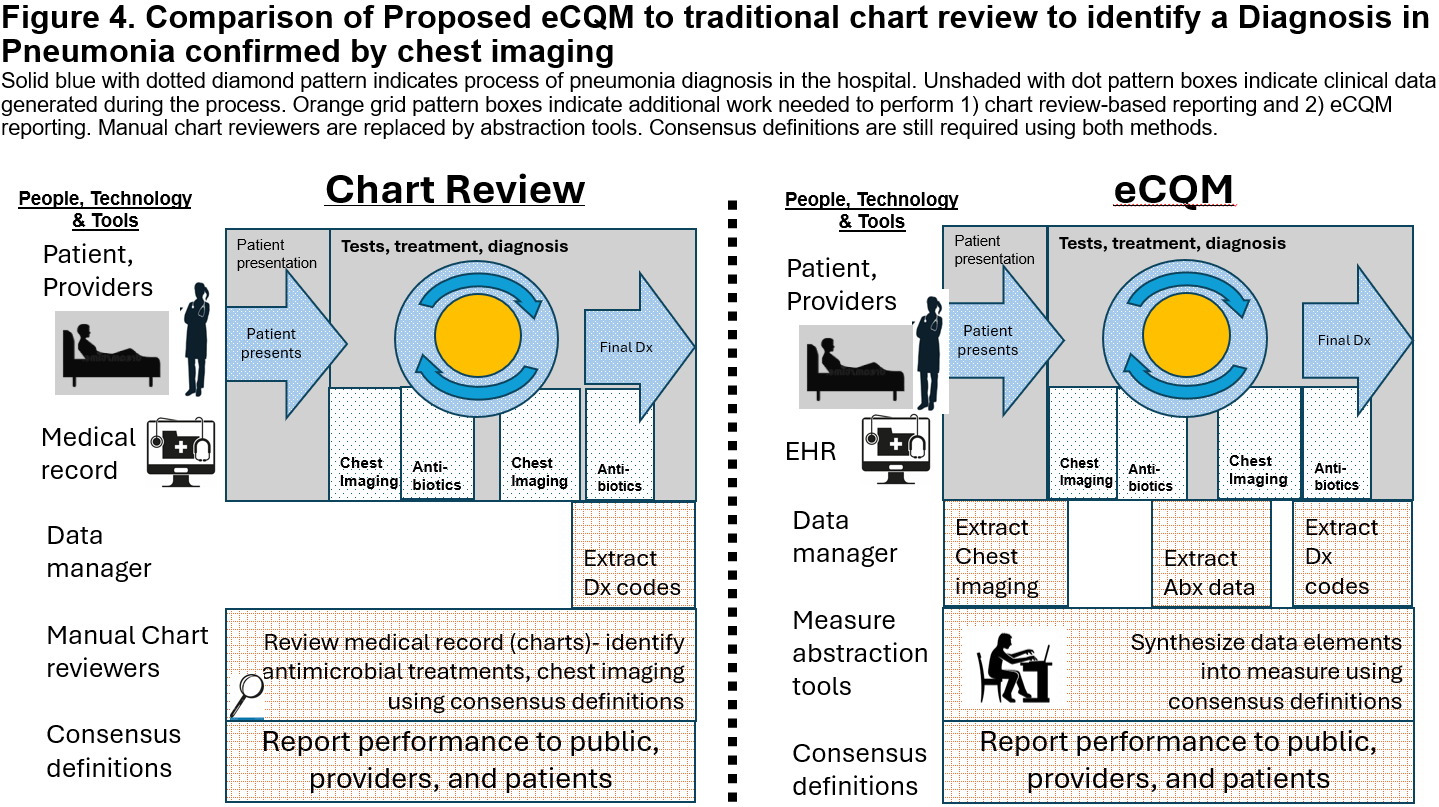
Each TEP was approximately an hour and a half long, with thirty minutes of orientation to rationale for the eCQM, current specification, and most current reliability and validity testing results. Following a presentation by the eCQM development team, moderators unaffiliated with the measure development 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. Table 6 describes the expertise and roles of

Table 6. Technical expert panel (TEP) participants during development and final measure specification phases.

|  |  |
| --- | --- |
| **Development phase TEP participants (N=4 Panels)** | |
| **Role** | **Expertise** |
| University of Utah Chief Informatics Officer | Informatics, EHR interoperability |
| Emergency department physician | Clinician |
| Intermountain Healthcare Chief of Pulmonary and Critical Care Medicine | Pneumonia |
| Chief of Health Policy & Quality Informatics, Baylor College of Medicine and Houston VA | Diagnosis |
| Infectious disease physician, measure developer | Infectious disease, healthcare safety |
| Patient (Univ. of Utah) | Patient experience: chronic lung disease with multiple episodes of pneumonia |
| Patient (Univ. of Utah) | Patient experience: interstitial lung disease, prior longstanding history of mis-diagnosis |
| **Final measure specification TEP (N=1 Panel)** | |
| Pharmacist and Acute Care Lead, Office of Antibiotic Stewardship, Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention (CDC) | Infectious disease, public health, antimicrobial stewardship |
| Infectious Diseases Society of America (IDSA) clinical practice guideline committee co-chair; Chief of Division of Infectious Diseases, Univ. of Louisville | Pneumonia and influenza |
| American College of Chest Physicians (CHEST), American Thoracic Society clinical practice guideline committee member, Associate Program Director of Division of Pulmonary and Critical Care Medicine, UT San Antonio | Pneumonia clinical practice guideline author, pulmonary & infectious disease |
| American Thoracic Society (ATS) quality improvement committee chair, Chief of Division of Pulmonary, Critical Care and Sleep Medicine, Univ. Of Connecticut | Pulmonary and critical care medicine, pneumonia care guidelines author |
| Hospital representative, Univ. of Chicago  Society of Hospital Medicine member | Hospital medicine, health outcomes research, pneumonia, quality measurement |
| Patient (Univ. of Utah) | Patient experience |
| Patient (Univ. of Utah) | Patient experience |

**3.1 – Feasibility Assessment**

Diagram of people, tools, tasks, and technologies required to implement the proposed eCQM (Figure 4). All data are collected as part of the normal clinical workflow but the data extraction, processing, and submission (indicated with orange grid-patterned boxes) represent additional steps required to format and submit the clinical data.



**3.1 Feasibility Assessment –**

Feasibility of Implementing Varying Measure Definition Stringency Levels within VA and University of Utah Health Systems:

* Measure definitions represent all possible combinations of the numerator (N), denominator (D), and antimicrobial timing (A) definitions shown below. Definition 1 (shaded column) represents the measure as specified and submitted for endorsement.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| \*\* | **Definitions** | **Def1**  **(N1, D1, A1)** | **Def2**  **(N1, D1, A2)** | **Def3**  **(N1, D1, A3)** | **Def4 (N1, D2, A1)** | **Def5 (N1, D2, A2)** | **Def6 (N1, D2, A3)** | **Def7**  **(N2, D1, A1)** | **Def8 (N2, D1, A2)** | **Def9 (N2, D1, A3)** | **Def10 (N2, D2, A1)** | **Def11 (N2, D2, A2)** | **Def12 (N2, D2, A3)** |
| Numerator (N) | Positive CXR/CT for pneumonia, where:  N1: + image any time  N2: + image 24 hrs prior to admission, hospital day 1 or 2 | N1 | N1 | N1 | N1 | N1 | N1 | N2 | N2 | N2 | N2 | N2 | N2 |
| Denominator:  Diagnosis (D)  AND  Antibiotic (A) | Diagnosed with pneumonia, where:  D1: inpatient ICD code for pneumonia at any position (primary, principal, or secondary)  D2: inpt ICD code for PNA in principal or primary OR secondary ICD code for PNA with primary/principal code for sepsis or respiratory failure | D1 | D1 | D1 | D2 | D2 | D2 | D1 | D1 | D1 | D2 | D2 | D2 |
| Treated for pneumonia (abx), where:  A1: antibiotics at any time  A2: receipt of antibiotics w/in 24 h prior and 72 h after initial presentation  A3: receipt of antibiotics w/in 24 h prior and 48 h after initial presentation | A1 | A2 | A3 | A1 | A2 | A3 | A1 | A2 | A3 | A1 | A2 | A3 |

\*\*Cell intentionally left blank.

**Variation in Measure Performance Score by Definition (VA Health Care System).** Shaded row indicates performance on measure if specified using that definition.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **\*\*** | **\*\*** | **Def1** | **Def2** | **Def3** | **Def4** | **Def5** | **Def6** | **Def7** | **Def8** | **Def9** | **Def10** | **Def11** | **Def12** |
| **Denominator Total** | Denominator count (A & D) | (A1 & D1): 89,767 | 85,686 | 84,379 | 75,181 | 74,487 | 74,017 | 89,767 | 85,686 | 84,379 | 75,181 | 74,487 | 74,017 |
| Dx count (D only) | D1: 90,822 | 90,822 | 90,822 | 75,927 | 75,927 | 75,927 | 90,822 | 90,822 | 90,822 | 75,927 | 75,927 | 75,927 |
| Abx count (A only) | A1: 1,135,184 | 1,068,524 | 1,039,365 | 1,135,184 | 1,068,524 | 1,039,365 | 1,135,184 | 1,068,524 | 1,039,365 | 1,135,184 | 1,068,524 | 1,039,365 |
| **Numerator Total** | Numerator count (N) | (A1 AND D1 AND N1): 80,820 | 76,950 | 75,738 | 67,626 | 66,970 | 66,537 | 74,362 | 72,747 | 71,973 | 65,548 | 65,190 | 64,877 |
| **Score performance for definition** | **% of patients dx'd with pneumonia and treated w abx who have a positive chest image** | **90.03%** | **89.80** | **89.76** | **89.95** | **89.91** | **89.89** | **82.84** | **84.90** | **85.30** | **87.19** | **87.52** | **87.65** |

\*\*Cells intentionally left blank.

**Variation in Measure Performance Score by Definition (University of Utah).** Shaded row indicates performance on measure if specified using that definition.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **\*\*** | \*\* | **Def1** | **Def2** | **Def3** | **Def4** | **Def5** | **Def6** | **Def7** | **Def8** | **Def9** | **Def10** | **Def11** | **Def12** |
| **Denominator Total** | Denominator (A&D) count | 3,030 | 3,028 | 3,026 | 2,204 | 2,203 | 2,203 | 3,030 | 3,028 | 3,026 | 2,204 | 2,203 | 2,203 |
| Dx count (D only) | 5,231 | 5,231 | 5,231 | 2,745 | 2,745 | 2,745 | 5,231 | 5,231 | 5,231 | 2,745 | 2,745 | 2,745 |
| Abx count (A only) | 16,960 | 16,951 | 16,931 | 16,960 | 16,951 | 16,931 | 16,960 | 16,951 | 16,931 | 16,960 | 16,951 | 16,931 |
| **Numerator Total** | Numerator count (N) | 2,785 | 2,783 | 2,782 | 2,041 | 2,040 | 2,040 | 2,698 | 2,696 | 2,695 | 2,000 | 1,999 | 1,999 |
| **Score Performance for definition** | **% of pts dx'd with pneumonia & treated w/ abx who have a positive chest image** | **91.91%** | **91.91** | **91.94** | **92.60** | **92.60** | **92.60** | **89.04** | **89.04** | **89.06** | **90.74** | **90.74** | **90.74** |

**3.1 Feasibility Assessment –**

Table 7a. Computing resources required for the optional NLP algorithm for chest imaging.

|  |  |  |  |
| --- | --- | --- | --- |
| **Implementation Site** | **Number of documents to process** | **Approximate processing time per document (seconds)** | **Estimated total compute time (hours)** |
| VA (5-year full dataset) | 2,240,391 | 0.2 | 124 |
| University of Utah (5-year dataset) | 43,035 | 0.04 | 0.48 |
| University of Michigan (6.5 years) | 132,345 | 0.04 | 1.5 |

Table 7b. Human resource requirements for implementing NLP algorithm for chest imaging in a new system unfamiliar with NLP (University of Michigan).

|  |  |
| --- | --- |
| **Human Resource Needs** | **Time (human hours to achieve)** |
| Training time for new data scientist to install NLP | 15 |
| Implementation of NLP | 2 |
| Validation and Customization within new EHR | 6\*\* |
|  | (in progress at U. Michigan) |

\*Windows installation package has been revised to reduce this time for future implementation.

\*\*Estimate (work in progress). Hours include 1-hour clinical time for validation/error analysis and 5 hours data analyst for customization.

**4.1.1 Data Used for Testing**

Figure 5a. VA testing cohort development (N= 89,767 hospitalizations meeting denominator definition).

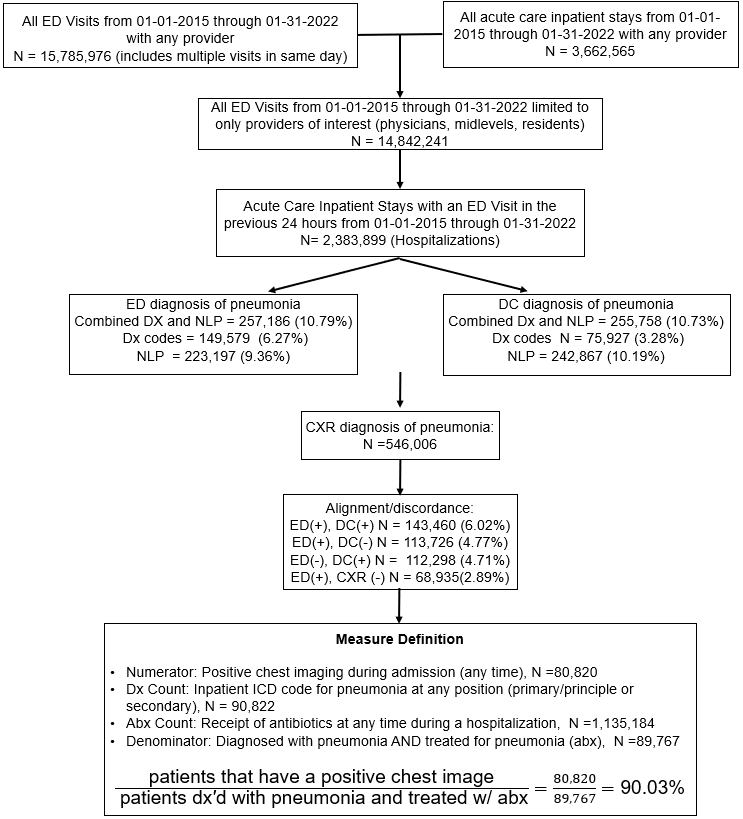


Figure 5b. University of Utah testing cohort development (N=3,030 hospitalizations meeting denominator definition).

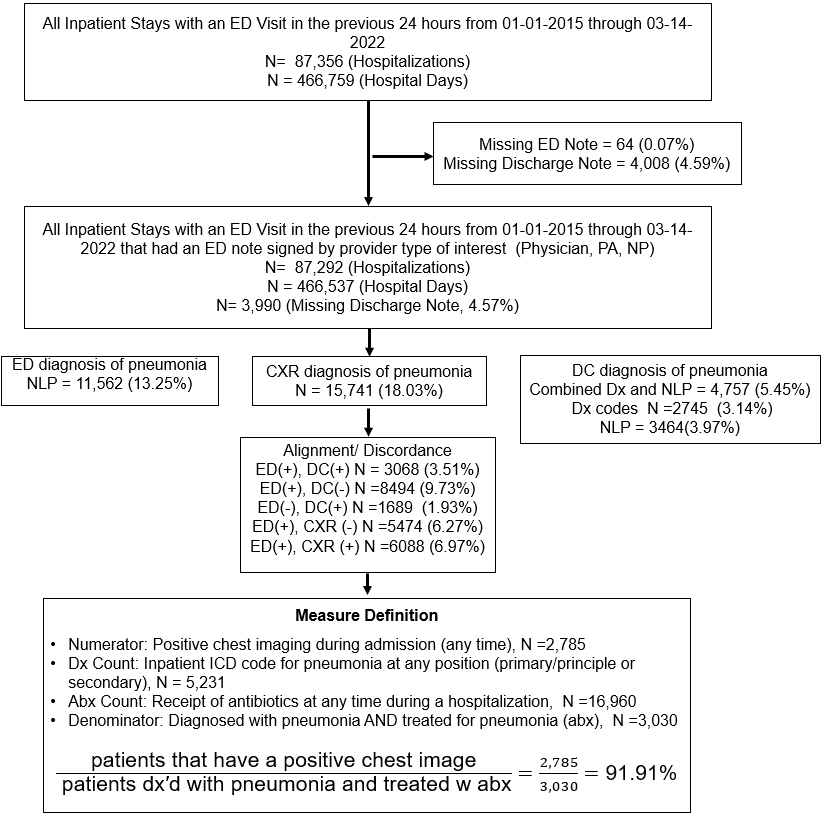
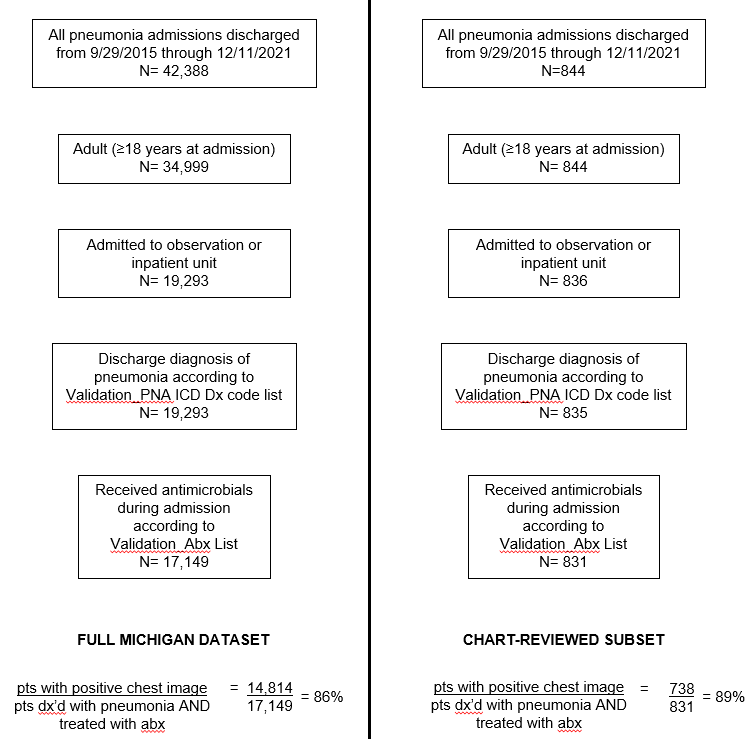


Figure 5c. University of Michigan cohort development (N= 17,149 patients meeting denominator definition for full University of Michigan dataset; N= 831 patients meeting denominator definition for chart-reviewed subset).



**4.1.2 Differences in Data**

Table 8. Types and levels of reliability and validity testing conducted on the proposed chest imaging-confirmed pneumonia diagnosis eCQM and descriptions of datasets used for testing.

| **Testing type and level** | **Data / Sample Population Used** |
| --- | --- |
| *Encounter Level Reliability*  -ICD-10 discharge diagnosis of pneumonia  -receipt of antimicrobial  -chest imaging report consistent with pneumonia | University of Utah & VA:  Initial population were patients with a diagnosis of pneumonia (any position) in emergency department or during inpatient admission, as identified by ICD-10 or NLP algorithm. Sample reviewed for data element reliability was a random sample of 50 hospitalizations per healthcare system with a discharge diagnosis of pneumonia from the full time period of the dataset. |
| *Accountable Entity Level Reliability* | VA Only:  All acute care inpatient stays between 1/1/2015 and 1/31/2022 with an emergency department visit in the preceding 24 hours with either an admission or discharge diagnosis of pneumonia identified by ICD-10 or NLP diagnosis algorithm. Reliability testing restricted to a single year of most recent data within cohort (1/1/2021-12/31/2021), which had 8,253 hospitalizations across 100 acute care facilities. |
| *Encounter Level Validity*  -Prevalence, sensitivity, and positive predictive value (PPV) of:  -ICD-10 discharge diagnosis of pneumonia  -receipt of antimicrobial  -chest imaging report consistent with   pneumonia | University of Utah & VA:   1. Initial population were patients with a diagnosis of pneumonia (any position) in emergency department or during inpatient admission, as identified by ICD-10 or NLP algorithm. Sample reviewed for data element reliability was a random sample of 50 hospitalizations per healthcare system with a discharge diagnosis of pneumonia from the full time period of the dataset. 2. Among hospitalizations with a discharge diagnosis of pneumonia (as identified by ICD code alone) and treatment with antimicrobials, a weighted random sample of 26 hospitalizations per site with chest imaging results consistent with pneumonia and 26 hospitalizations per site without chest imaging consistent with pneumonia. [only PPV calculated for this sample] 3. To validate the chest imaging result consistent with pneumonia data element: Within the denominator population dataset for each site (discharge diagnosis of pneumonia and treated with antimicrobials), chest imaging results (see “Validation\_ChestImagingCPTCodes” spreadsheet for list of imaging codes used) were text searched for strings found in the “Validation\_Chest Imaging” spreadsheet. The NLP algorithm was run on the body of records with any of these search terms identified. The NLP algorithm may be found at: <https://github.com/abchapman93/medspacy_pneumonia>   University of Michigan:  Validity testing was performed in the chart-reviewed 831 patient dataset where patients had a discharge diagnosis of pneumonia and received antimicrobials. |
| *Encounter Level Validity*  -PPV of the eCQM measure performance | University of Utah and VA:  Among hospitalizations with a discharge diagnosis of pneumonia (as identified by ICD code alone) and treatment with antimicrobials, a weighted random sample of 25 hospitalizations per site with chest imaging results consistent with pneumonia and 26 hospitalizations per site without chest imaging consistent with pneumonia. [only PPV calculated for this sample] |
| *Evidence of Performance Gap & Assessment of Measure being “Topped Out”* | VA Only:  All acute care inpatient stays between 1/1/2015 and 1/31/2022 with an emergency department visit in the preceding 24 hours with either an admission or discharge diagnosis of pneumonia identified by ICD-10 or NLP diagnosis algorithm. Testing sample was restricted to a single year of most recent data within cohort (1/1/2021-12/31/2021), which had 8,253 hospitalizations across 100 acute care facilities. |
| *Empiric Validity/ Associations with outcomes* | VA Only:  Initial population were patients with a diagnosis of pneumonia (any position) in emergency department or during inpatient admission, as identified by ICD-10 or NLP algorithm who received antimicrobial treatment between 2015 and January 2022 (n=89,767), the full VA dataset. |

**4.1.3 Characteristics of Measured Entities**

Table 9. Characteristics of health systems the proposed eCQM was tested in.

|  |  |  |  |
| --- | --- | --- | --- |
| **Hospital Characteristic** | **University of Utah**  **N=1 hospital**  **(3,030 hospitalizations)** | **VA Health System**  **N=103 hospitals**  **(89,767 hospitalizations)** | **University of Michigan**  **N=1 hospital**  **(17,149 hospitalizations)** |
| Academic Hospital | 1 (100%) | n/a | 1 (100%) |
| Location  Urban  Rural | 1 (100%)  0 (0%) | 84 (81.6%) 19 (18.4%) | 1 (100%)  0 (0%) |
| Hospital Type  State  Federal | 1 (100%)  0 (0%) | 0 (0%)  103 (100%) | 1 (100%)  0 (0%) |
| Bed size  ≤50  51-100  101-200  >200 | 0  0  0  1 (100%)\* | 24 (23.3%)  38 (36.9%)  38 (36.9%)  3 (2.9%) | 0  0  0  1 (100%)\* |
| Annual patient census | 29,450 | 485,579 | 49,730 |

\*Note: University of Utah Hospital has 524 beds; the University of Michigan: 1,043 beds.

**4.1.4 Characteristics of Units of the Eligible Population** \*

Table 10. Characteristics of VA patients in the testing dataset (N=89,767).\*

| **Variable** | **All Patients (pneumonia discharge diagnosis + antimicrobial treatment)**  **N=89,767 (%)** | **Chest Imaging Positive N=80,820 (%)** | **Chest Imaging Negative**  **N=8,947 (%)** |
| --- | --- | --- | --- |
| Age, median (interquartile range) | 72 (65, 81) | 72 (65, 81) | 72 (65, 82) |
| Sex  Male  Female | 86,295 (96)  3,472 (3.9) | 77,714 (96)  3,106 (3.8) | 8,581 (96)  366 (4.1) |
| Race  White  Non-white | 68,974 (77)  20,793 (23) | 61,925 (77)  18,895 (23) | 7,049 (79)  1,898 (21) |
| Rural residence | 27,322 (30) | 24,421 (30) | 2,901 (32) |
| BMI, median (SD); 18.3% missing | 27 (23, 32) | 27 (23, 31) | 28 (24, 33) |
| Married | 42,224 (47) | 38,030 (47) | 4,194 (47) |
| Housing instability | 8,601 (9.6) | 7,714 (9.5) | 887 (9.9) |
| Nursing home (previous 90 days) | 2,864 (3.2) | 2,583 (3.2) | 281 (3.1) |
| Hospitalization (previous 90 days) | 22,507 (25) | 20,424 (25) | 2,083 (23) |
| Comorbidities (missing for 13.6% of patients)  Cerebrovascular disease  Congestive heart failure  COPD  Dementia  Diabetes without complications  Diabetes with complications  History of myocardial infarction  Peripheral vascular disease | 13,588 (15)  25,142 (28)  40,914 (46)  7,081 (8.0)  17,210 (19)  10,314 (12)  7,712 (8.7)  13,396 (15) | 12,164 (15)  22,682 (28)  36,938 (46)  6,305 (7.9)  15,494 (19)  9,332 (12)  6,913 (8.7)  12,139 (15) | 1,424 (16)  2,460 (28)  3,976 (45)  776 (8.8)  1,716 (19)  982 (11)  799 (9.0)  1,257 (14) |
| Charlson comorbidity score, median (IQR) | 3 (1, 5) | 3 (1, 5) | 2 (1, 5) |
| Risk Score: ePSI <90  Missing | 30,835 (35)  2,582 (2.9) | 27,441 (35)  2312 (2.9) | 3,394 (39)  270 (3.0) |
| Hospital complexity  High  Medium/low  Unknown | 78,568 (88)  11,034 (12)  165 | 70,744 (88)  9,932 (12)  144 | 7,824 (88)  1,102 (12)  21 |
| Positive blood or respiratory cultures (missing for 7.4% of patients) | 13,825 (17) | 12,569 (17) | 1,256 (16) |
| Any chest imaging (CT or x-ray) | 81,508 (91) | 75,937 (94) | 5,571 (62) |
| Positive chest image | 70,916 (79) | 70,887 (88) | 29 (0.3) |
| Received antibiotic during hospitalization | 89,767 (100) | 80,820 (100) | 8,947 (100) |
| Diuretic during hospitalization | 33,904 (38) | 31,202 (39) | 2,702 (30) |
| Steroid during hospitalization | 35,269 (39) | 31,905 (39) | 3,364 (38) |
| Antiviral in first 24 hours | 5,392 (6.0) | 4,712 (5.8) | 680 (7.6) |
| Median days of antibiotic treatment (IQR) | 5 (4, 8) | 5 (4, 8) | 4 (3, 6) |
| Time in hours to first antibiotic, median (IQR) | 3 (2, 6) | 3 (2, 6) | 3 (2, 6) |
| Admitted to the ICU | 10,520 (12) | 9,791 (12) | 729 (8.1) |
| Length of stay in days, median (IQR) | 5 (3, 9) | 5 (3, 9) | 4 (3, 6) |
| Inpatient mortality | 4,577 (5.1) | 4,429 (5.5) | 148 (1.7) |
| 7-day mortality | 1,847 (2.1) | 1,718 (2.1) | 129 (1.4) |
| 30-day mortality | 8,107 (9.0) | 7,678 (9.5) | 429 (4.8) |
| 1-year mortality | 27,505 (31) | 25,386 (31) | 2,119 (24) |
| 30-day readmission  Missing | 13,766 (16)  4,577 | 12,509 (16)  4,429 | 1,257 (14)  148 |
| Discharge Location  Home  Hospice  Nursing home  Other hospital  Other  Unknown | 73,056 (81.4)  509 (0.6)  7,816 (8.7)  1,406 (1.6)  1,273 (1.4)  5,707 (6.4) | 65,300 (80.8)  475 (0.6)  7,185 (8.9)  1,320 (1.6)  1,125 (1.4)  5,415 (6.7) | 7,756 (86.7)  34 (0.4)  631 (7.1)  86 (1.0)  148 (1.7)  292 (3.3) |
| 30-day mortality after discharge | 10,178 (11) | 9,679 (12) | 499 (5.6) |
| 60-day mortality after discharge | 13,328 (15) | 12,589 (16) | 739 (8.3) |
| 1-year mortality after discharge | 27,721 (31) | 25,587 (32) | 2,134 (24) |
| 60-day readmission  Missing | 20,271 (24)  4,577 | 18,336 (24)  18,336 | 1,935 (22)  148 |
| 90-day readmission  Missing | 24,625 (29)  4,577 | 22,243 (29)  4,429 | 2,382 (27)  148 |
| Nursing home admission within 30 days | 1,418 (1.7) | 1,281 (1.7) | 137 (1.6) |

\*Any that do not sum to 100% is due to missing data for a particular patient characteristic

Table 11. Characteristics of University of Utah patients in the testing dataset (N=3,030).

| **Characteristics** | **All Patients (pneumonia discharge diagnosis + antimicrobial treatment)**  **N = 3,030 (%)** | **Chest Imaging Positive Patients**  **N =2,785 (%)** | **Chest Imaging Negative Patients**  **N= 245 (%)** |
| --- | --- | --- | --- |
| Unique admitting ED providers, by type  Physician  Midlevel  Total unique providers | 130 (77.8)  37 (22.2)  167 | 126 (78.3)  35 (21.7)  161 | 73 (80.2)  18 (19.8)  91 |
| Age (years), mean (SD) | 59.4 (17.9) | 59.4 (17.8) | 60.0 (19) |
| Sex  Female  Male  Unknown | 1,328 (43.8)  1,701 (56.1)  1 (0.03%) | 1,214 (43.6)  1,570 (56.4)  1 (0.04) | 114 (46.5)  131 (53.5)  0 (0) |
| Race  White or Caucasian  Black or African American  Asian  American Indian or Alaska Native  Native Hawaiian/Other Pacific Islander  Other/Choose not to disclose/unknown | 2,345 (77.4)  70 (2.3)  82 (2.7)  53 (1.8)  77 (2.5)  403 (13.3) | 2,150 (77.2)  67 (2.4)  72 (2.6)  52 (1.9)  71(2.6)  373 (13.4) | 195 (79.6)  \*removed others due to minimum cell counts |
| Marital status  Divorced  Legally separated  Life partner/domestic partner  Married  Single  Other/unknown  Widowed | 302 (10.0)  40 (1.3)  81(2.7)  1,410 (46.5)  799 (26.4)  63 (2.1)  335 (11.1) | 278 (10.0)  39 (1.4)  78 (2.8)  1,291 (46.4)  736 (26.4)  60 (2.2)  303 (10.9) | 24 (9.8)  <10  <10  119 (48.6)  63(25.7)  <10  32 (13.1) |
| Time to antibiotics (hrs), median (IQR) | 2.6 (1.6-4.1) | 2.6 (1.6-4.0) | 3.1 (1.8-5.5) |
| CT scan obtained | 1,396 (46.1) | 1,342 (48.2) | 54 (22.0) |
| Length of stay in days, median (IQR) | 5 (3-8) | 5 (3-9) | 4 (3-6) |
| 28-day readmission | 387 (12.8) | 357(12.8) | 30 (12.2) |
| Discharge location  Home  Death  Hospice  Skilled nursing facility  Other | 2,140(70.6)  240 (7.9)  85 (2.8)  472 (15.6)  93 (3.1) | 1,954 (70.2)  231(8.3)  76 (2.7)  437 (15.7)  87 (3.1) | 186 (75.9)  <10  <10  35 (14.3)  <10 |
| 30-day mortality | 108 (3.6) | 102 (3.7) | <10 |
| Inpatient mortality | 253 (8.4) | 244 (8.8) | <10 |

Table 12. Characteristics of chart-reviewed testing dataset, University of Michigan (N= 831).

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **All Patients (pneumonia discharge diagnosis + antimicrobial treatment)**  **N=831 (%)** | **Chest Imaging Positive**  **N=738 (%)** | **Chest Imaging Negative**  **N=93 (%)** |
| Age, median (IQR) | 67 (55-77) | 67 (55-77) | 70 (57-78) |
| Sex  Male  Female | 473 (56.9)  358 (43.1) | 421 (57)  317 (43) | 52 (55.9)  41 (44.1) |
| White race | 683 (82.2) | 605 (82.0) | 78 (83.9) |
| Married | 354 (42.6) | 316 (42.8) | 38 (40.9) |
| Previous hospitalization (prior 90 days) | 285 (34.3) | 249 (33.7) | 36 (38.7) |
| History of myocardial infarction  Congestive Heart Failure  COPD  Peripheral vascular disease  Cerebrovascular Disease  Dementia  Diabetes without complications  Diabetes with complications | 105 (12.6)  225 (27.1)  359 (43.2)  88 (10.6)  41 (4.9)  55 (6.6)  130 (15.6)  142 (17.1) | 88 (11.9)  **194 (26.3)**  319 (43.2)  76 (10.3)  38 (5.1)  49 (6.6)  123 (16.7)  120 (16.3) | 17 (18.3)  **31 (33.3)**  40 (43.0)  12 (12.9)  3 (3.2)  6 (6.5)  7 (7.5)  22 (23.7) |
| Charlson Comorbidity Index | 2 (1-3) | 2 (1-3) | 2 ()1-3 |
| Positive blood or respiratory cultures | 163 (19.6) | 142 (19.2) | 21 (22.6) |
| Chest imaging obtained | 830 (99.9) | 738 (100) | 92 (98.9) |
| Chest tomography obtained (53.3% missing) | 388 (46.7) | **369 (50.0)** | **19 (20%)** |
| Positive chest image | 738 (88.8) | 738 (100) | 0 (0) |
| Antibiotic in first 24 hours | 822 (98.9) | 732 (99.2) | 90 (96.8) |
| Antibiotic during hospital | 831 (100) | 738 (100) | 93 (100) |
| Inpatient antibiotic days, median (IQR) | 4 (3-5) | 4 (3-5) | 3 (2-4) |
| Time to first antibiotic, median (IQR) hours | 2.9 (1.9-4.6) | 2.9 (1.8-4.6) | 3.2 (2.2-5.1) |
| Length of Stay (days), IQR | 5 (3-6) | 5 (3-7) | 4 (3-5) |
| In the ICU during first 48 hours | 4 (0.5) | 4 (0.5) | 0 (0) |
| Inpatient Death | **7 (0.8)** | **7 (0.9)** | **0 (0)** |
| Discharge diagnosis code for pneumonia | 831 (100) | 738 (100) | 93 (100) |
| Discharge Location  Home  Home Health  SNF  Other | 566 (68.1)  145 (17.4)  96 (11.6)  17 (2.0) | 498 (67)  132 (18)  86 (12)  15 (2) | 68 (73.1)  13 (14.0)  10 (10.8)  2 (2.2) |
| 28-day Readmission | 112 (13.5) | 99 (13) | 13 (14) |
| 28-day ED Visit | 56 (6.7) | 51 (7) | 5 (5.4) |

**4.2.2 – Method(s) of Reliability Testing**

Description of Cohen’s kappa methodology:

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 a statistical estimate of the extent of agreement between clinician reviewers beyond that expected by chance. Cohen’s kappa was calculated as:

κ = Pr(observed agreement) – Pr(chance agreement)

1 – Pr(chance agreement)

Where Pr(observed agreement) = number of agreements between clinician reviewers / number of charts reviewed

and, Pr(chance agreement) = ( [(total number of charts reviewer 1 marked the data element as present \* total number of charts reviewer 2 marked the data element as present) / total number of charts reviewed] + [(total number of charts reviewer 1 did not consider the data element present) \* (total number of charts reviewer 2 did not consider the data element present) / total number of charts reviewed] ) / total number of charts reviewed

**5.1 Contributions Towards Advancing Health Equity**

Table 21. Proposed measure performance by patient demographic characteristics among patients from VA system.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Full VA Cohort**  **(2015-2022)** | | | **Validation VA Cohort (2021)** | | **Full UU Cohort**  **(2015-2022)** | |  |
| **Patient Characteristics** | **N** | **Measure Score % (SE)** | **p-value** | **N** | **Measure Score % (SE)** | **N** | **Measure Score %** | **p-value** |
| Rural residence  Rural  Non-rural | 27,322  62,445 | 89.4 (0.12)  90.3 (0.19) | <0.001 | 2,521  5,732 | 91.7 (0.55)  93.0 (0.34) | \*\* | |  |
| Older Adult  Over 80 years  80 years and under | 23,192  66,575 | 89.5 (0.20)  90.2 (0.12) | 0.002 | 1,794  6,459 | 91.1 (0.67)  93.0 (0.32) | 381  2,649 | 91.1  92.0 | .04 |
| Age  Over 65 years  65 years and under | 67,060  22707 | 90.0 (0.12)  90.1 (0.20) | 0.86 | 6,413  1,840 | 92.4 (0.33)  93.1 (0.59) | 1,207  1,823 | 91.2  92.4 | .008 |
| Race  White  Non-white | 68,974  20,793 | 89.8 (0.12)  90.9 (0.20) | <0.001 | 6,172  2,081 | 92.1 (0.34)  94.0 (0.52) | 2,345  685 | 91.7  92.7 | .34 |
| Sex  Male  Female | 86,295  3,472 | 90.1 (0.10)  89.5 (0.52) | 0.26 | 7,900  353 | 92.6 (0.29)  92.1 (1.4) | 1,701  1,329 | 92.3  91.4 | .005 |
| Complexity of hospital care sought  High complexity  Medium/Low complexity  Uncategorized | 78,568  11,034  165 | 90.0 (0.11)  90.0 (0.29)  87.3 (2.6) | 0.94 | 7,284  895  74 | 92.6 (0.31)  92.2 (0.90)  95.9 (2.3) | \*\* | |  |

\*\*Cells intentionally left blank.

**6.2.1 Actions of Measured Entities to Improve Performance**

Table 22. TEP panelist feedback on importance of the Chest Imaging-Confirmed Diagnosis of Pneumonia eCQM.

| Illustrative Quote | TEP representative | Theme |
| --- | --- | --- |
| *“I took a little trip to the emergency department not too long ago and had pneumonia, and [after reviewing information on diagnostic quality in pneumonia] I think I asked better questions, just to the doctors… I think that I was able to ask different questions in a different way, about my treatment and care and stuff like that.”* | Patient | Usability |
| *“I think the measurement system has to be easy to access, easy to read and easy to follow up and add on to. I am not sure if those are possible things to ask for since the hospital computer system is such a massive source of information already.”* | Patient | Usability – patient communication |
| *Obviously we know that there's always education that can happen, but education by itself is not usually a good way to change behavior. So you need a systems approach. I think there are plenty of ways that this measure be usable. I think the greatest threat to usability would be potential skepticism… If there's acceptance of the validity of the measure, I think it's eminently usable.* | Clinician, pneumonia quality leader | Usability |
| *“I think that if we can automate the measure or build clinical decision support that make doing the right thing to do the easy thing to do as well, I think it will have a significant impact on patient care. If it's not integrated well into the provider workflow in an efficient manner, the impact will be limited.”* | Clinician, informatics leader | Usability – clinical decision support |
| *“…this could really complement…our CAP work and…I think there 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 this really, the you pursuing [chest imaging confirmation] is really helpful and can benefit…the quality measure for CAP in general.”* | CDC director of antimicrobial stewardship | Usability – foundation for other eCQMs |
| *“I think a consistent definition that can use the electronic medical record to identify patients so that there is no not as many differences as we see now. For example, 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.”*** | Hospitalist and pneumonia quality expert | Usability – foundation for other eCQMs |
| *“I’m really intrigued by the concept of diagnostic discordance as a continuous improvement measure, not only in pneumonia but across diseases…And so, I think in general there’s a need for a paradigm shift in the way that providers receive information about diagnosis and diagnostic discordance. And that as a continuous improvement approach.”* | Clinician, chief informatics officer / operations leader | Usability - continuous quality improvement |
| *“I I'm sure that we need to separate as you mentioned community-acquired pneumonia… and I would concentrate all the efforts initially in the possibility of Community acquired pneumonia… if you are able to develop a measurement that you can move across hospitals and across electronic medical records, one thing that is critical here that he has been always the challenge is reading the radiologist notes extracting from the radiologist note…”* | Clinician | Usability – foundation to other eCQMs |

The transcript summaries and/or methods for the TEPs follow. Please note, the first four TEPs were for an earlier measure specification focused on the emergency department. ***Only the last TEP directly addresses the current measure specification.***

Technical Expert Panel

(TEP): Measuring and

Improving Diagnostic

Excellence in Pneumonia

May 17, 2021

Researchers: Barbara Jones, MD, MDCI

Facilitated by: Brieanne Witte, BS

Scribed by: Naomi Flake, BS

Report prepared by: Naomi Flake, BS & Brieanne Witte, BS

Objectives

Technical Expert Panel: Measuring and Improving Diagnostic Excellence in Pneumonia

Researcher: Barbara Jones, MD, MDCI

Research Project Background

Pneumonia is commonly misdiagnosed by providers, as diagnosis includes a level of uncertainty. Providers identify pneumonia from non-sensitive or specific health conditions. They then test their assumptions referencing previous research and experiences that require quick accurate responses. The final diagnosis is often identified using a combination of test results and delayed information after an initial patient visit. Once pneumonia is diagnosed, it can be difficult to change the treatment. Often, Emergency Department physicians do not receive feedback about patients’ testing results and diagnosis.

The research team is seeking to improve the diagnostic process for both patients and emergency providers by:

* Collecting data from the VA and University of Utah electronic health records
* Comparing Emergency Department provider diagnoses of pneumonia with the diagnoses of the discharging hospital provider
* Comparing Emergency Department provider diagnoses of pneumonia with chest x-rays (radiographic diagnoses)
* Measuring conflicting pneumonia diagnoses and returning patient information to clinic providers

To make this project successful, input is needed from experts like you. As a Technical Expert Panel, we will review and discuss the research plan. Feedback from the panel will be used to support and improve the project from design to implementation.

Demographics

The Technical Expert Panel members consisted of six participants, representing the following roles and organizations:

* Chief Informatics Officer, University of Utah
* Emergency Department Physician, University of Utah
* Chief, Pulmonary & Critical Care, Intermountain Healthcare
* Chief, Health Policy & Quality Informatics, Baylor College of Medicine and Houston VAMC
* Patient, University of Utah (2)

Agenda

|  |  |
| --- | --- |
| 11:00 a.m. - 11:15 a.m. | **Welcome and Introductions**   * Name * Organization/Role * What made you decide to join this panel? * What do you expect to gain from this experience? |
| 11:15 a.m. - 11:25 a.m. | **Group Norms & Expectations** |
| 11:25 a.m. - 11:50 a.m. | **Diagnosis Experience Questions**  • What processes have you experienced when receiving a medical diagnosis?  o What contributes to a patient being satisfied? o What contributes to a patient being dissatisfied? |
| 11:50 a.m. - 12:05 p.m. | **Project Presentation** |
| 12:05 p.m. - 12:50 p.m. | **Presentation Questions**   * What information did you find interesting in this presentation/report? * Would this measure be useful for providers to recognize diagnostic patterns in pneumonia? o Why or why not? * Would reading a report about this measure be helpful to you personally?   o Why or why not?   * What impact do you think this measure will have on patient care? * What other diagnosis do you think this type of measure could be useful for? * What do you anticipate will occur while developing and using this report tool? |
| 12:50 p.m. - 1:00 p.m. | **Wrap-up and Next Steps** |

Transcript Summary

1. **What processes have you experienced when receiving a medical diagnosis?**

a. Patient had predominantly positive experiences with receiving a medical diagnosis

i. Has become familiar with the process of medical diagnosis through experience

b. Patient spent three years going to various doctors to figure out what was wrong

i. Each doctor gave patient the same diagnosis

None provided an x-ray or pulse oximeter test

None of the interventions were successful

* + 1. Went to a lung specialist and patient received a CAT scan, pulse oximeter test,

etc.

* + - 1. Given the correct diagnosis
      2. Spent three years trying to get doctors to listen to them about how they were feeling

c. Patient went to Urgent Care

* + - * 1. Provider found focal crackles
        2. Provider suggested pneumonia diagnosis without additional tests
        3. Patient prescribed antibiotics and recovered

d. Most common misdiagnosis is pneumonia

* + - * 1. 1 in 20 adults are going to be misdiagnosed each year
        2. Patient had pneumonia but did not receive diagnosis in time (I.e. unexpectedly hospitalized and then diagnosed)

e. Most common reasons for misdiagnosis (see: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/16565 40):

* + - * 1. Ignoring patient concern
        2. Clinician-patient interaction

Providers not listening to patients

Not reviewing patient medical history iii. Labelling condition as pneumonia when it is not the correct diagnosis (i.e. over-diagnosing)

1. Missed in primary care and emergency setting iv. Most of the time, red flag signs and symptoms are not recognized when they should have been

v. Only labelled as a misdiagnosis when there is clear evidence that clinician should have done something different

f. Challenges to diagnostic process:

i. Difficult diagnosis

Diagnosis changes over time

Under and over-diagnosis

a. New study suggests practitioners overestimate the probability of disease, including pneumonia, before and after testing (see:

https://jamanetwork.com/journals/jamainternalmedic ine/fullarticle/2778364)

* + - * 1. Patient history and exams are not empirically clear
        2. Patients may have multiple health conditions

1. As a provider, choose to focus on which is most important to treat

iv. Limited information and limited time to diagnose correctly as an

Emergency Room doctor

g. Successful diagnosis process (i.e. diagnosis is clear and both patient / provider knows what is going on):

* + - * 1. Patient history is suggestive (e.g. cough, fever)
        2. Physical exam (e.g. focal crackles)
        3. Chest x-ray
        4. Provider returns results to patient
        5. Provider and patient discuss treatment plan
        6. Provider and patient discuss expectations
      1. Electronic clinical decision support tool:
      2. Reviews all patients in network
         1. Uses algorithm to calculate likelihood of pneumonia
         2. Requires x-ray finding of lung opacity for diagnosis
         3. Displays to Emergency Department (ED) physician when likelihood of pneumonia diagnosis reaches 40%

1. Produces 50/50 probability of pneumonia diagnosis 2. Successes:

ED physicians diagnose on their own

When likelihood of pneumonia is not very high, think more broadly then they would originally

3. Issues:

Delay for interpreting correctly long-winded free text reports from radiologists

There is no gold standard for pneumonia; sometimes have to hold professional panel to decide

v. Developing technology for real-time lung x-ray findings

Interested to see if ED physicians will use it more, as it is more accurate and timely

Need to learn more about ED physician interest (i.e. if they want to use a tool to improve diagnosis accuracy)

1. **What contributes to a patient being satisfied?** 
   * 1. Understanding, empathy, and compassion drive patient satisfaction
     2. Ensuring patients can get in to be seen, especially on an urgent basis \*
        + 1. Never having to wait for an appointment
          2. Never having to wait to see provider
     3. Provider that is trusted, open, and willing to listen to complaints \*
     4. Bi-directional communication between patient and provider

i. Not sugar-coating information to patient

* + 1. Patient understanding and inclusion in decision-making
    2. Clear discharge information
    3. Healthcare team working together to ensure patient is doing okay (i.e.

having someone in the patient’s corner)

* + 1. Patient follow-up
    2. Ensure patient is seen again and doing better

i. Personalized care

* + - 1. Treating patient like a human being and not just a number on a paper \*
      2. Completing patient health tests as seamlessly as possible in the   
          shortest amount of time
         1. Checking oxygen levels
         2. Getting blood drawn

k. Addressing all concerns in one visit so patient does not have to return for additional tests

i. Cardiologist came in during the same appointment with lung specialist to talk with patient

l. Ensuring a warm hand-off between patient and other departments \*

i. Improves patient experience because they do not have to reexplain everything

1. **What contributes to a patient being dissatisfied?**

a. Provider talking down to patient (e.g. “I am the doctor and you are the dummy”)

i. Sugar-coating information for the patient

b. Downplaying or not accepting a patient’s experience about how they are feeling \*

i. Trivializing how a patient feels (e.g. “you don’t really feel that way”)

c. Miscommunication

i. Healthcare team tells the patient different information

After month-long pulmonary experience, patient sent home with various papers with checked and unchecked items that were confusing

Had to call pulmonologist to clarify medications being terminated, doubled, changed, etc.

1. **What information did you find interesting in this presentation/report?**

a. Quality score cards

* + - * 1. There is a lot of data on how well they are received
        2. Some are poorly designed and do not provide good mechanisms for medical education
    1. Mortality is different between discordant diagnoses
    2. Morbidity and Mortality Conference (M&M)
       - 1. Shifting towards a discussion about what everyone could have done better to avoid negative patient outcomes in the future
         2. System issues create bad outcomes
         3. Want to have real-time information to help providers

d. Approach ER doctors to provide this tool to help them reach their goals

(i.e. support intrinsic motivation)

i. Do not want the tool to come across as watching or monitoring doctors’ tasks

e. Good to frame this type of measure as feedback for self-learning rather than good or bad performance

i. Some of the obvious 'missed' cases with bad outcomes may need different methods of delivering non-punitive feedback (see:

https://www.jointcommissionjournal.com/article/S1553-

7250(20)30234-8/abstract)

f. It is great

* + - * 1. More inclusive to everyone getting the same material and information
        2. More streamlined and connected
        3. Everything is in one place and available to everyone

g. Sometimes a misdiagnosis feels like the patient’s fault

i. Patient does not look like or present as typical transplant patient, or the right questions are not asked to get to the correct diagnoses

1. Loss to follow-up

1. **Would this measure be useful for providers to recognize diagnostic patterns in pneumonia? Why or why not?**

a. Useful for intrinsic desire for provider to do the right thing

i. ER doctors want to work well and have rapport with their colleagues

b. Always a risk that these measures – although developed for improvement

– eventually are used for accountability

i. Need to ensure that does not happen with this measure

c. Past experience has shown that private providers and groups were not interested in using a pneumonia diagnostic tool

* + - * 1. Tool most commonly did not reach outside of ER directors
        2. Most likely shared with groups but not discussed further

d. Biggest challenge:

i. Provider adoption

Busy

Need to consider how to deal with different personalities and group psychology

Quality scorecards used only if incentivized to use them

e. Include informaticists

* + - * 1. Make sure tweaked in correct direction to make sure it meets needs of physicians embedded within the emergency department
        2. Previous experience has taught that tools without informaticists resulted in poor adoption compared to group with informaticist
        3. Think about how to present the data

f. Think about filtering the tool and making it a learning experience instead of one-time feedback opportunity

1. **Would reading a report about this measure be helpful to you personally? Why or why not?** 
   * 1. Data presentation needs to be informative and easy to digest in quick   
         timeframe
     2. Visually informative
        + 1. Provide 1 – 2 items provider can work on
          2. If there is a lot of data, provider may not know what to make of it

c. False positive / false negatives:

i. Have a tool that shows the main issue with each of those wings returned to the provider (e.g. missing lab value)

1. **What impact do you think this measure will have on patient care?**

a. Patients come back to the hospital for similar things

i. Allows opening for possible conversation between providers and patients during second visit

1. Potential to improve care by providing diagnosis that will help patient get better
   1. Not about being right or wrong, but about improving patient

outcomes

1. When patient first diagnosed, nobody wanted to talk about it
   1. It is difficult to find someone who will listen, look at all the pieces, and put it together
   2. In full support if this tool makes it easier for doctors to diagnose or treat patients
2. May help providers communicate better with their patients (e.g. sharing   
    uncertainty with a patient to help them know a condition may develop   
    further)
   1. Would rather have a doctor be honest and say they do not know everything about patient’s circumstances
   2. May help improve confidence in ER physicians

1. Previous experience with ER provider refusing to communicate with patient’s primary care provider makes patient hesitant to go back to the ER for care iii. Need to communicate uncertainty better to patients in the correct way (see: https://academic.oup.com/intqhc/article/30/1/2/4791877)

iv. Have used this concept in calibration (i.e. accuracy and confidence in diagnosis)

* + 1. When provider gets the patient diagnosis right, they are confident
    2. When provider is not sure of diagnosis, their confidence should fall
    3. If better calibrated, provider will do less under and overdiagnosis
       1. To become better calibrated, provider needs feedback on own performance
       2. Timing of calibration is important

i. Providers most receptive to learning with the next patient that is like their last patient

* + 1. Upstream feedback (see:

https://qualitysafety.bmj.com/content/early/2021/05/10/bmj qs-2020-012464)

* + 1. Calibration and feedback (see: https://jamanetwork.com/journals/jama/fullarticle/27247928 7968976)

Additional Questions

1. **How are you identifying best practices in those better at diagnosing pneumonia?** 
   1. Triangulate with this discordance measures and others to identify best   
       practice
   2. Validity: 
      1. Seeing whether this measure has any relationship to any other clinical measures that are important
   3. Setting-specific 
      1. May favor timely treatment over diagnosis in busy emergency room
      2. In smaller less busy settings, may have more time to review and make a diagnosis prior to treatment
   4. Main goal: see if the tool can help doctors educate themselves and   
       come up with continuing medical education that helps improve   
       diagnostic skills

1. **Do you think that more discordance measures could help providers have more   
    humility (i.e. to look things up or to ask for help?)** 
   1. Yes

1. **Are you planning on giving all discrepancies or just a few?** 
   1. Discrepancies are on a gray scale (i.e. not black and white)
   2. Feedback has to be unbiased, accurate, and timely
      1. Quarterly report of consistent list of all pneumonia patients linked to

EPIC ii. Normative feedback:

* + - 1. Score card shows how aggregate statistics compare to their peers
      2. Does not tap into intrinsic motivation iii. Task feedback:
      3. More important than normative feedback
      4. Want to promote task feedback in the tool (e.g. getting a gold   
          coin after completing a task)

iv. Standard feedback:

* + - 1. Expect certain diagnostic accuracy
      2. Currently not created for diagnostic tool

Post-Meeting Survey Responses

1. **What processes have you experienced when receiving a medical diagnosis?** 
   * 1. “Physical exam, Pulmonary function tests, Labs, radiographs, CT seem to be the first series of items performed.” (Lacee)
     2. “Explaining in clear terms, helping me to understand.” (Megan)
     3. “I generally have been fortunate to be pretty healthy. When I have received a medical diagnosis, it has been during one-on-one conversations in the physician in the exam room.“ (Maia)
2. **What contributes to a patient being satisfied?** 
   * 1. “I think being listened to and acknowledged to how you are feeling, explaining why different procedures may be ordered. Being a part of decisions being made on your behalf.” (Lacee)
     2. “Clear expectations and plan for follow up/antibiotics etc.” (Megan)
     3. “Having enough information to make an informed decision.”
        + 1. “Feeling that they were part of the shared decision-making   
              process and that conversation occurred with empathy and   
              respect.”
          2. “A team-based approach. Where there is confidence that the   
              transition from one specialty to the next is seamless.” (Maia)

1. **What contributes to a patient being dissatisfied?** 
   * 1. “Lack of communications or empathy from the physician. Being left out of   
         conversations that directly pertain to your healthcare plan. Feeling like the   
         doctor isn't really concerned about how you progress after you leave.”   
         (Lacee)
     2. “Unclear expectations or no clear plan, no clear diagnosis.” (Megan)
     3. “A breakdown of any of the above (listed in question 2).” (Maia)
2. **What information did you find interesting in the presentation/report?**

a. “I really enjoyed that there is such a broad range of participants, and also

I thought it was very interesting that both of the patients are from Agriculture backgrounds. I really enjoyed listening to the ER doctor s perspectives because she really is the first line when the patient goes through the emergency system. I very much enjoyed the different viewpoints and conversations, even though I was really nervous!” (Lacee)

* 1. “So interesting that the mortality of those patients who had the incorrect diagnosis was higher.” (Megan)
  2. “I most sound interesting the perspective of patients. We make a lot of assumptions as providers on what a patient would want or not want based on our deeper understanding of medicine as well as our own bias based on what we would want his patients.” (Maia)

1. **Would this measure be useful for providers to recognize diagnostic patterns in pneumonia? Why or why not?** 
   1. “I think the measurement system has to be easy to access, easy to read and easy to follow up and add on to. I am not sure if those are possible things to ask for since the hospital computer system is such a massive source of information already?” (Lacee)
   2. “Yes, to help learn from their practice patterns.” (Megan)
   3. “Yes I think it would be useful.” (Maia)

1. **Would reading a report about this measure be helpful to you personally? Why or why not?** 
   1. “I definitely think it would be educational and informative, but I am not sure what I personally could do with the information after.” (Lacee)
   2. “Yes, but the report should be useful and easy to use, clear with visuals so that the provider can digest it quickly.” (Megan)
   3. “I think reading a report would be helpful, but as a busy provider there are so many areas of medicine to keep up with. With best intentions we try to read as many reports as possible, but realistically speaking this is not the case.” (Maia)

1. **What impact do you think this measure will have on patient care?** 
   1. “I think it would be invaluable to be able to follow a pattern on the patient specifically, But I do realize not all patients present the same every time they seek medical services.” (Lacee)
   2. “Hopefully to decrease bad outcomes (mortality and taking a long time to come to the right diagnosis).” (Megan)
   3. “I think that if we can automate the measure or build clinical decision support that make doing the right thing to do the easy thing to do as well, I think it will have a significant impact on patient care. If it's not integrated well into the provider workflow in an efficient manner, the impact will be limited.” (Maia)
2. **What other diagnosis do you think this type of measure could be useful for?** 
   1. “Seizure patients, chronic pain? Obviously, it won’t be able to measure everything but it could help those patients with repeated admissions to

ER.” (Lacee)

* 1. “Chest pain, heart failure, asthma.” (Megan)
  2. “Not sure.” (Maia)

1. **What do you anticipate will occur while developing and using this report tool?** 
   1. “I think the panel with see avenues that may work really well and others may look good on paper but fail to thrive in real world use.” (Lacee)
   2. “Providers might not want to be "watched" too closely or told that they are doing a bad job.” (Megan)
   3. “I imagine there be a lot of back-and-forth between various stakeholders including IT/informatics, patient advocacy groups, clinicians, and care teams.” (Maia)

1. **What do you believe are the most important contributions the TEP has made to the research project?** 
   1. “Patient views are not always taken into account in many studies, and I think it is amazing you are reaching out for patient participation.” (Lacee)
   2. “Hearing voices from many sources, patients, providers, research.”

(Megan)

* 1. “Clarifying plans for measurement.” (Hardeep)
  2. “My hope is that having the diverse contributions of multiple stakeholders has made the research project of increased value to the organization and various stakeholders including our patients.” (Maia)

1. **What has been most important in helping the TEP function well?** 
   1. “Having the moderator to keep thing running smoothly and keeping things flowing well.” (Lacee)
   2. “Great organization and clear instructions.” (Megan)
   3. “Cohesive team.” (Hardeep)
   4. “Clear communications and administrative support to ensure each participant was informed and present.” (Maia)
2. **What has made TEP participation most worthwhile for you personally or professionally?** 
   1. “I feel like after my nerves calm down a bit being called upon. I may be able to contribute with more precise information.” (Lacee)
   2. “Learning from Dr. Jones and patients.” (Megan)
   3. “Not sure yet. Looking forward to ongoing participation.” (Maia)

1. **What suggestions do you have for improving the TEP meeting?** 
   * 1. “Keep up the great work!” (Lacee)
     2. “Earlier time to finalize meeting time.” (Megan)
     3. “I think best time to capture thoughts is at the meeting.” (Hardeep)
     4. “No suggestions at this time.” (Maia)

Measuring and Improving Diagnostic Excellence in Pneumonia

Technical Expert Panel Session 2

Date: September 27, 2021

Presentation: Barbara Jones

Facilitator: Teresa Taft

Organizer: Lindsay Carter

Report prepared by: Teresa Taft

Attendees

* David Classen, ED & Epidemiology physician, University of Utah
* Maia Hightower, Chief Informatics Officer, University of Utah
* Megan Fix, Emergency Department Physician, University of Utah
* Nathan Dean, Chief, Pulmonary & Critical Care, Intermountain Healthcare
* Hardeep Singh, Chief, Health Policy & Quality Informatics, Baylor College of Medicine and Houston VAMC
* Lacee Sims, Patient, University of Utah

Study team in attendance

* Alec Chapman
* Barbara Jones
* Charlene Weir
* Jian Ying
* Jori Butler
* Lindsey Carpenter
* Michaela Nevers
* Teresa Taft

Background

Pneumonia is commonly misdiagnosed by providers, as pneumonia diagnosis includes a level of uncertainty. Providers identify suspected pneumonia from symptom presentation, referencing previous research and experiences. Notably, the action - outcome feedback learning loop is often not complete for Emergency Department (ED) physicians as they infrequently receive feedback about patients’ testing results and final diagnosis.

Objectives

We are seeking to improve the diagnostic process for the benefit of both patients and emergency providers by providing ED physicians with outcome feedback for patients they have diagnosed with pneumonia. To make this project successful, input from expert physicians, researchers and patients is needed. Participants were asked to share their experiences and insights to provide guidance on our measure development and refinement, and alignment with practice settings.

TEP #2 Meeting Summary

***Complexity in determining a diagnostic discordance.*** Participants discussed the complexity of determining a pneumonia diagnosis from the medical record, including consideration of community vs. hospital acquired pneumonia and if radiographic evidence should be the gold standard for this work. Dr. Jones clarified, that for this study, any mention of pneumonia without an exclusion is considered a diagnosis, which is admittedly a broad definition.

***NLP methods.*** A suggestion was made to develop customized NLP rules for specific settings where a note is written in the patient’s hospital journey. Dr. Jones clarified that the NLP is optimized to capture pneumonia rule-out or diagnosis mentions at 3-points in time (ED, Radiology, Discharge) and that the intention is for the less focused NLP tool to have good performance across hospitals.

***Measures to consider beyond inter-rater reliability.*** Important areas to focus on in evaluating the feedback measure report may include: 1) measure accuracy, 2) diagnostic accuracy, 3) under and over diagnosis, 4) downstream effect. In addition, it may be interesting to assess the effect on the level of care that patients receive. A warning and recommendation against including hospital readmission in the analysis was given.

***Relevance of the measure.*** Participants described the need for the feedback report to provide insight that is actionable. Identifying patterns of overdiagnosis leading to antibiotic overuse is an area of interest. A participant pointed out that physicians may be pushed toward more conservative pneumonia diagnosis, to improve their scores, if this report is rolled out. The research team discussed the possibility of looking at timing when the diagnosis changes in a future study. Clarification of the ‘Treatment’ column in the report was requested to identify the point in the hospital journey when the ‘treatment’ was given. This suggests a need to improve titling, i.e. ‘ED Treatment’. Adding a column to the report specifically indicating whether or not antibiotics were given was suggested. Currently the report has a column where this information is included with other treatment data. It was also suggested that an ideal, (though not currently feasible) version of this feedback report would provide diagnostic error attributions and best practice recommendations. A suggestion was made to focus on specific types of patients. This might be done through a narrow application of the feedback measure report, subgroup analysis, or by adding a filter feature which would allow providers to select the type of patient they are interested in.

***Intrinsic motivation to review feedback measures.*** There was concern that people may not be interested in the report if it feels like another ‘quality measure’ telling them they did something wrong. Without incentivization, the diagnosis report may not be used. Presenting the report at staff meetings was suggested, to overcome initial barriers to use.

***Planning ahead.*** A question regarding when the participant could imagine using the feedback measure report in their workflow would be valuable. Preliminary results of the user study will be shared at the next TEP meeting. In a future study, looking at the effect, on patients, of changing diagnosis could be interesting.

Quotations to support this summary and responses to the post-meeting survey can be viewed in their respective sections below.

TEP #2 Agenda

2:00 – 2:05pm ***Welcome***

2:05 – 2:10pm ***Reintroductions***

*15 seconds to share* something that you enjoyed this summer

*2:10 – 2:25* ***Presentation - Update of Progress***

*2:25 – 2:55* ***Chart review in UUH and* VA *system***

1. *Measures beyond inter-rater reliability & validity*
2. *What to look for in error analysis: distinguishing “misdiagnosis” from diagnostic discordance.*

*2:55 – 3:25* **Assessing provider interactions with the measure**

1. *Overview of self-theories and impact on performance*
2. *Efficiency, clarity and scope*
3. *Potential to affect learning & behavior*
4. *Intrinsic motivation*
5. *Sampling/recruiting approach (to include representation from diverse group of providers)*

*3:25 – 3:55* **Evaluating associations between the measure and clinical outcomes**

1. *Existing outcome measures that may be correlated*
2. *Additional relevant outcomes (clinical, or other)*
3. *Meaningfulness of correlation, Causal inference & Confounding issues*

3:55 – 4:00pm **Wrap-up**

1. REDCap questionnaire instructions
2. Next Meeting

TEP #2 Meeting Quotations

***Complexity in determining a diagnostic discordance***

* How is a diagnosis of pneumonia is determined for the measure/report?
* we are trying to look for any possible case, so if they mentioned pneumonia as a possible source of sepsis, then that would be a positive case. If pneumonia is mentioned, even in their MDM, that's positive. Now that can mean that it can be overly broad. If they don't clearly exclude pneumonia, then we keep that as a positive.
* Determine whether you require radiographic abnormality consistent with a radiographic pneumonia before diagnosing pneumonia… that's particularly important when it comes to conventional chest imaging versus CT scan because CT scan still has a specificity problem, but it's probably sensitive for pneumonia… You can kind of hone down the chart review to those who have radiographic disease If you define pneumonia as requiring radiographic changes, which is as you know, the way the guidelines have always done pretty much with the exception of a couple stray papers require the cohort to be selected with chest imaging.
* Another thought that had was that an issue might be that you have pneumonia onset after hospital admission that might have been a hospital acquired pneumonia or a ventilator associated pneumonia. And so when we've been looking at emergency department accuracy of pneumonia, I try to put myself in the ED physicians mind and say, “you know if there was no radiographic pneumonia at the time, then the patient didn't have pneumonia and if they develop pneumonia later on, although, of course, some people will have progressive changes on the radiograph and that are evident on subsequent films.
* I've never really been misdiagnosed, I always just get a change in what you get, you know, as it's morphed into something else. Like when all this started, they told me I had walking pneumonia. Then they told me I had lympho sarcoma. And of course, I had to go through every specific battery of tests to rule that one way or the other. Then I had to have some lymph node biopsies, which showed necrotizing granuloma and the only thing that came up with that results was Wegener's granulomatosis, which did not fit my age, my gender or my color. And then since then, it's changed to … chronic lung disease. Right now I'm dealing with cytomegalovirus, … I've always kind of had these little debt problems that go along with everything. And so they've said, Oh, you've got chronic bowel disease, and then it's … I don't really ever feel like I've been misdiagnosed. I just feel like it's an ever changing… Maybe, I have morphed throughout the last 14 years into a different animal used to be.

***NLP methods***

* Have you considered having multiple NLP data sets, say one in the emergency department, another when they're … the patient's journey at least in my mind is a different data set and hence the NLP labeling will be will be different and if you think about it as one NLP derived data set then it may lose that temporalness as well as decrease the accuracy depending on where we are in the patient journey.
  + this is a rule-based NLP … so there are kind of building blocks … But we have it set up so that for each note type, we're changing the way the words are used and the rules that it's around… and so it's set up, so that ultimately, one could design something pretty easily and modular, that would go to say, a progress note. So then on day five, we could actually have an NLP tool that is trained to be capturing that type of note, and actually be using these really similar concepts, but in really different ways so that it's tweaked, and, and trained to be optimized for that specific note. So we chose to optimize our NLP tools for these specific slices in time, because they're so standard, everyone gets an ER visit, and everyone, you know, er, document, and then everyone gets a discharge summary. And everyone hopefully, gets a chest imaging report… But trying to actually develop something in something really noisy, makes the tool a little bit more resilient to different systems. So that was our rationale. So we ended up using being kind of the opposite, where we, we tried to optimize this big program for many different note types in different, you know, slices of time, and then kind of sub trained these tools onto these three slices.

***Measures to consider beyond inter-rater reliability***

* 1st, Verifying that what your measure pulled out from the database is truly what you intended it to pull out. So, if you're looking at some discrepancy, or quota diagnosis, that's what it's showing, and you're trying to use NLP to help you do that
* 2nd Validity testing; you're trying to figure out what's the reason for this discordance so I'm thinking there's at least three or four reasons for all of these; one is it's just according an accuracy for some reason or the other second is involving diagnosis and there's no way you could have done anything different, 3rd it’s a quality problem and there're two kinds of quality problems. One is under-diagnosis, that means you miss pneumonia, or it's over diagnosis, 4th is downstream effects. If this was a truly a discordant diagnosis, then bad things can happen such as you know moderate morbidity, mortality, length of stay, etc/
* if there was a change in level of care, so if someone was admitted to the floor, then they went to ICU and vice versa. That's something that I like to look at.
* I would leave hospital readmission … referring to patients who were initially sent home who secondarily are admitted within a week. Hospital readmission is a complete can of worms that's more related to comorbid illness and processes of care post discharge than it is to any process of care in the emergency department

***Relevance of the measure***

* Under this the part that says treatment, what does that mean? Does that mean that's what the ED physician gave?
* I think reviewing the notes, which is the real time, like that specific patient example is helpful, but I also think that reviewing what is the best practice that I'm aiming for, so that I can understand the principle that I should be applying to the next patient, so that really is what makes it a learning health system or a learning feedback loop-- when tI can then say, “Oh, this is the best practice that I sort of misfired on”, and how if a similar patient comes to present, that I can then do better … you know, a visit similar to that sort of misdiagnosis, then I can apply the principles that I didn't apply the first time around to the next patient. So it's that combination of the specific information from the patient's chart, as well as the broader principle that I can learn from.
  + And I'm not sure whether you would want to do that, I think long term, that's really cool. But with this short term project of one year, I think maybe you just want to show them the outcomes,…
* I'm just wondering if the measure might be more useful for specific types of patients. And if you can focus on which ones I'm thinking. And I think the comment about well, if you have too many comorbidities, you know, you're more likely to get, you know, not only just to discordance, but many other things, right? --readmitted and all that things. We did a study where we just looked at the lower risk patients who got escalated from the inpatient unit to the ICU, to see if we can find any preventable adverse events. The The idea was from a global trigger tool from IHI that talks about escalation of care. And we said, well, let's just look at low risk escalation of care, because we know many of these patients who are, you know, old and have lots of comorbidities and you really can't find any signal, whether there was a safety issue or not, and I'm thinking for your work, could you maybe look for some of those factors, maybe it's more useful for people who have who are younger and have less cool mobility, maybe you shouldn't be doing this in patients who have pre existing lung disease and COPD, because, you know, they are battle-lands to begin with, and the x rays are maybe abnormal and things like that, and that'll help you get the value out better.
  + So thinking maybe these outcomes would be about the same but doing subgroup analyses of you know, maybe high risk patients low risk patients
* If you know, this data is valuable, you're giving it to clinicians, and let them go check it out. If there is no value in your data, if there is no signal in your data, if I look at 20 records like this, and I find nowhere that I can improve, or maybe one, that data is will not be very useful to me.
* So if I'm a clinician I want to know you know, what the data that you're showing me is useful for me so it's provided you get the data right … Are you going to sort of try to make a case that the data that you're going to push in front of the providers of, you know, while they were going to pull is valuable to them? Because I haven't seen that value or the proposition for that value yet.
* Aside from diagnostic discordance, one of the other things that would be important is chart review to review the treatment decisions. ED physicians, when they have been given feedback say, “Well, if I did the right thing, then, you know, did it really matter that I might have gotten the diagnosis wrong, but I, I did the right thing, so I may not have said that they had pneumonia, but I gave them antibiotics for something else. So that didn't really change their outcome.
* Does this measure enhance our understanding of how many of those patients that were over-diagnosed actually had a non-infectious disease at the end and therefore, we were giving them antibiotics that were unnecessary? That could be one thing that that could be really important. … Is it something that you could tease out? A lot of people get overdiagnosis pneumonia, and then they end up having a non-infectious disease and that might be something that's really different from having another infectious disease that might have really much pretty much the same treatment.
* I think that Ed physicians tend to view not prescribing antibiotics for a patient with infection as a much worse problem than prescribing the wrong antibiotics. And probably they are sued more if they miss and don't prescribe antibiotics for an infection. But you know, with Brandon Webb's work and others, if you give vancomycin and zosyn to patients who should be treated with or ceftriaxone their appear to be significantly worse outcomes with the broader spectrum antibiotics for multiple reasons… I think, Ed physician tend to over prescribe antibiotics and partly for the reason that I just said,
* If I was a really smart Ed physician, and I wanted to score well on your test, I would diagnose pneumonia only in a very straightforward low bar, productive sputum fever and elevated white cell count. And the iffy uncertain cases, that would shy away from and then as a percentage, the accuracy of diagnosis would be much higher.
* [Researcher] And one thing we were thinking of is down the road, you know, we could do something where the question is, how much time did it take the diagnosis to change? So if you have a change in diagnosis, one thing is that, you know, we actually are very vulnerable to that initial diagnosis. So we kind of attached emotionally to that. And then that can carry with the patient for a long time before people start to question it or realize that it's not the right diagnosis.
  + [TEP physician] somebody can be doing the exact right thing to do in the emergency department. And then the situation evolves, you get more information, the right thing to do say on day a hospital day two, and then the right thing to do on discharge being but yet have discordance in that diagnosis. But having a more, of course we haven't gotten there in science, but a dynamic quality measure. So that actually assesses over time, that changes over time. And it would be helpful for me as a clinician, because what I decided in the emergency department is probably going to be different than day five have a hospital stay.
* oftentimes there's over treatment with antibiotics-- should there be another column with recommended treatment … I'm just wondering, for someone who wants to really improve with this tool, there should be an area where it's like suggestions or something like that… are specific, evidence based **best practice** that I should be adhering to? And where did I go wrong, so to speak? Or did I even go wrong? … There’s like a richness that is lacking.

***Intrinsic motivation to review feedback measures***

* The deal breaker for this is, if it doesn't have any consequences in what people do.
* As we've rolled out a pneumonia tool across Intermountain the physicians, and actually, the group, psychology seems to vary quite a bit, some physicians are really quite interested and want to get better and look at their data and others just want to know if they're in trouble and if not go out mountain biking, or skiing, or whatever. So there's really quite a bit of variability in how people would view this tool and how much they would use it unless there was a reward or a penalty.
* I think it's best when, yes, we give them information, and they actually discuss it in a group meeting. I think that that is a great motivator. But some of the groups really don't even care enough to want to do that. Even if I provided the information. They just, again, want to get their scheduling done and get on with things.
* sometimes when you're just given information, it may seem like someone's telling you that you did something wrong, that can be a turn off. … the wording and the messaging about what this information is, is super important because if someone is just thinking that these people are doing this pneumonia study, they're going to be telling me that I'm doing errors, or I'm doing something wrong, that's going to turn off people from trying to get the information. And so for me personally, like for our group, net, you talked about the group at IMC and that some people just don't care, I think that you have to message it to them appropriately so that they do care …I think people would be much more willing to engage with you, and engage with your idea of like, this is all about making things better for the patient and teaching us how to be better, that would be much more well received than someone just getting this dashboard and saying, Oh, I only have 75% accuracy. … this is a lot of data that is going to need to have a little bit more explanation and messaging from your team, for people to accept it, Because I don't think people intrinsically don't want to learn, I think that if they are feeling attacked, then they're going to just turn off.
* When will I do this? Is it a part of my, you know, daily work? Or is this something I'm supposed to do at nights and weekends when I'm not in the clinic, for instance, which you know, I've done primary care or is, so that's why we have to make this as a part of somebody's workflow. If it's useful. Once it's once you prove that this is useful, then I think people will start looking for it in seek out that, but you're still going to need a lot of intrinsic motivation.
* what, if you … went to an organized group, … their weekly or monthly meetings, and show them the results of this, both in total and then each individual one, you could privately show their results? So you build it into their quality improvement meetings, or nightly meetings?
* I had developed quality scorecards in the past and believe me as much as we would hope that that Providers look at them and glean over them, they really just look at what's red, they usually will look at see what's green with what they did well, though this provider like 90%, sensitivity, Okay, I'm gonna give myself a pat on the back, because 90% there, okay, I like that. And then look at the one red spot and try to figure out, this is where the growth mindset, folks anyway, that looked at the scorecard and try to figure out what did I miss? And how can I apply it to a similar patient like that in the future? What evidence-based approach did I miss?

***Planning ahead***

* Incorporate a question in the user study about how, when, or where would you embed review of the report into your existing workflow.
* think if we maybe get more funding down the road to do a second round of this, maybe we would do something where we do patient interviews to really think about what the actual emotional consequences of the changes
* **Share some results of user testing at next TEP**

TEP #2 Post-Meeting Survey Responses

Participants were asked to complete an anonymous 16-question survey. Four TEP members responded to the survey, two of whom appear to be patient members. Although one of the patient members was not in attendance at the meeting, they were provided with a recording of the TEP meeting to review prior to taking the survey.

1

Does our measure of diagnostic discordance comparing admission diagnosis with radiographic and discharge diagnoses seem to be scientifically sound?

1. Yes

2. Yes

3. NA

4. yes, but there are some patients with multiple diagnoses for admit from the ED as well as multiple discharge dx. we should make sure to include them.

2

In what situations would this measure be useful?

1. In every situation

2. Respiratory patients

3. NA

4. reviewing cases for ED providers

3

In our measure we look at diagnostic discordance - when the admission diagnosis differs from the chest imaging (radiographic) and discharge diagnoses. What other types of discordance do you think we should we be including in our feedback tool, to help physicians get better at diagnosing?

1. Accurate patient history

2. Patient recurrences or previous experiences. Many chronic lung disease rarely present the same every time

3. Missed opportunities

4. treatment discordance, opportunity for improvement (treatment, avoidance of treatment, final micro or virology)

4

What treatment patterns could we include?

1. Don't broad treat, be specific. Correct diagnosis through proper testing

2. Previous admission reports?

3. NA

4. antibiotics, lack of antibiotics, steroids, increased level of care within 14 hours (e.g. ICU transfer)

5

In addition to final chest x-ray reports, what key diagnostic information might be missing at the time of Emergency Room Pneumonia diagnosis?

1. Correct and timely history

2. Prior history

3. history, exam, labs

4.micro, virology, blood cultures, CT reports

6

What is the meaning of diagnostic discordance to you?

1. Having as much information as possible (i.e. history, labs, x-ray, listen to patient )

2. A missing piece between admit and discharge

3. evolution of diagnosis or coding issue or missed opportunity in making a correct or timely diagnosis

4. difference between admit diagnosis and discharge diagnosis

7

When, in provider workflow, would it be most useful to receive this report?

1. As soon as possible after admit

2. Yes I believe it would

3. need protected time not when seeing patients

4. during faculty development time, quarterly?

8

Would you expect providers to be intrinsically motivated to use it in the pursuit of excellence in diagnosis, or would they only use this report if it was required?

1. I would hope the former, but fear it may be the latter. In the meeting I heard a lot about value added information or it may not be worth a provider's time.

2. If it’s easy to use and access I think more physicians will use it and share their experiences with purgers

3. mix of both

4. yes! I sure would

9

Are there other existing outcome measures that may be correlated with our measure?

1. NA

2. I am sure there will be some

3. NA

4.learning opportunities

10

What additional relevant patient outcomes should we consider displaying (clinical, or other)?

1. History, same outcomes from other admits.

2. We all want a good result and sometimes there are many ways to achieve this goal.

3. NA

4. death, LOS, increased level of care (ICU), discharge on O1

11

Do you believe there is a causal relationship between diagnostic discordance and patient outcomes or experience?  
  
If yes, through what mechanisms?

1. Yes. In my situation I was diagnosed incorrectly for 2 years even though I had a history of lung problems. I knew something was wrong but could not get a doctor to do an xray.

2. Yea, I think you can have a break in communication with patients not being forthcoming or having their health care in a different system that isn’t readily available.

3. NA

4. not always, need to figure out the cause, which would involve

12

What information did you find interesting in the presentation/report?

1. I also belive this study has to be of value or doctors will not use it.

2. There is a lot of work going into this and I hope that it works for physicians and patients both

3. NA

4. great work on this!

13

Do you expect that the measure/report will improve diagnostic accuracy or patient care in the Emergency Department?

1. I do if it is a standard that all providers will use to diagnose and treat.

2. I sure hope so

3. Could if associated with feedback

4. not sure, it might, would be really interesting to find out

14

What has made TEP participation most worthwhile for you personally?

1. Very interesting hearing all opinions. Nice to have a voice in the process and also a captive audience willing to listen.

2. I know my doctors work hard on my behalf and it’s nice to know that y’all on this board are real people working to the highest ability to help patients. IE me ;)

3. NA

4. getting to hear the perspectives of the patients and other panelists

15

What do you believe are the most important contributions the TEP has made to the research project?

1. Wide variety of providers willing to contribute, listen, learn and implement.

2. Different points of view, thinking on things maybe not thought of before

3. NA

4. thinking about helping providers learn how to better care for pneumonia patients.

16

What suggestions do you have for improving the TEP meeting?

1. An editorial on how to use meeting app, technologically challenged.

2. Keep up the good work! I think for as many people involved there has been participation.

3. Skip the survey-- elicit these at the meeting

4. its been great so far.

Recap of 1st TEP meeting

* Panel members value hearing the perspectives and experiences of this diverse panel.
* Patients value providers who listen, have compassion, provide clear information, personalized care and warm handoffs in addition to not having to wait.
* A diagnosis should only be labelled a misdiagnosis when there is clear evidence that something different should have been done. Diagnosis evolves over time.
* Challenges in making pneumonia diagnosis
  + Providers overestimate the probability of disease
  + Patient medical history is not clear
  + Co-existing problems to address.
  + Limited patient time & information availability in Emergency Department (ED)
  + There is no gold standard for pneumonia-- sometimes have to hold professional panel to decide
* Insights & Perceptions of Diagnostic Measure Report
  + Useful for improving provider calibration. (“Should I be confident in my diagnosis?”)
  + Should help providers reach their goals - Will the report make it easier to diagnose or treat patients?
  + Frame as feedback for self-learning, not a quality score. Providers don’t want to hear that they are doing a bad job.
  + The report should help providers identify their practice patterns.
  + Want all of the information in one place
  + Must be easy & quick to digest
  + Make it available to everyone
  + Should support conversations between colleagues
  + Feedback should be timely – most relevant to look at management of prior patients when seeing a similar patient.
  + The feedback must be integrated into the workflow.
  + What important clinical measures does this measure have a correlation with?

Moore PNA Dx Feedback Measure

TEP #3 Meeting Summary

Feb 22, 2022

**Attendees**

* David Classen, Infectious Disease Doctor, University of Utah
* Barbara Jones, Principle Investigator
* Maia Hightower, Chief Informatics, Officer Internist, University of Utah
* Megan Fix, Emergency Department Physician, University of Utah
* Hardeep Singh, Chief, Health Policy & Quality Informatics, Baylor College of Medicine and Houston VAMC
* Heath Young, Patient, Blackfoot Idaho

Study team in attendance

* Alec Chapman
* Barbara Jones
* Liz Rutter
* Jian Ying
* Jorie Butler
* Lindsey Carpenter
* Teresa Taft
* Peter Tabor

**Study background**

Context

As a patient moves through the hospital system, they may receive a diagnosis at one point, and the physicians don't really get feedback about whether that diagnosis is ultimately correct. So, among the patients who are hospitalized from the emergency department, the information evolves and then there's a final discharge diagnosis at the end of their hospitalization.

Looking at this discordance as well as discordance with a chest X ray diagnosis of pneumonia, we developed a measure that we call diagnostic discordance. The numerator is when those diagnosis are not the same and the denominator is all patients that are hospitalized from the emergency department with a diagnosis of pneumonia noted in their record at any point during the hospitalization. First, we count diagnosis codes across a patient's hospital stay, but that’s not very sensitive, so we extract the terms of pneumonia from clinical notes using natural language processing tools that we adapted for emergency department notes, chest imaging reports, and discharge summaries.

We developed a natural language processing tool in the fall, first in the VA system, which is a large system with lots of clinical notes that have lots of variation; and then we tested and refined it in the VA system. Then we tested it, off the shelf, meaning without customizing at all in the university of Utah Health system, and we then had a customization step.

Precision and sensitivity of the NLP tool

For emergency department notes the NLP detects mid 80s percent of pneumonia cases from the notes. For extracting pneumonia diagnosis from the radiology report, we trained it to be highly sensitive, because we're using it to verify clinical diagnoses that are positive for pneumonia, but it was not as important to be precise, meaning we wanted it to be highly sensitive --not to miss a diagnosis that was actually truly present. And then for discharge summaries, we tried to go for balance. The NLP tool’s precision degraded in the university data when applied to discharge summaries. (But the ability to extract the pneumonia diagnosis from the chest imaging or the CAT scan was pretty high without customization. And with minimal customization, we are pretty high precision as well. However, NLP captured more cases than the diagnosis codes for both the initial and the final dx.

Pneumonia dx in the VA

In analyzing data from 140 VA emergency departments across the nation, we found 2.2 million emergency department visits with hospitalization, from. About 10% of all of those hospitalizations had an emergency department diagnosis of pneumonia and bout 22% of those hospitalizations had a chest imaging report that had something that looked like a positive pneumonia and about 10% of those patients had a final discharge diagnosis of pneumonia. The 10% on the initial dx and the final dx are not the same. One third of the cases (133,000) had both an initial diagnosis and a discharge diagnosis of pneumonia. (true positive) One fourth of the cases (97,000) had an initial diagnosis in the emergency department for pneumonia, but did not have a discharge diagnosis of pneumonia (false positive). One fourth of the cases (104,000) had a discharge diagnosis of pneumonia, but did not have an initial diagnosis of pneumonia (false negative). And so that translates to a diagnostic positive predictive value of 40% and a sensitivity of 58%.

Possible reason for variation in precision and sensitivity across settings

We perceived that the structure of the clinical note at the University of Utah’s 2 emergency departments is very different from VA notes. People at University of Utah Hospital left the bulk of their medical diagnoses in their medical decision-making section.

Study objectives

1. We hope to determine what kind of feedback measure may help providers improve the accuracy of their diagnosis at the bedside. And so, we're looking at discordances, when diagnoses at the beginning and end of hospitalization are not the same.
2. Measure how much customization work is required to adapt an NLP tool that works really well, to extract diagnoses from clinical texts in one setting (the VA), for use in a different setting.

**TEP panel** **results**

There was a general consensus that creating a diagnostic discordance measure is cutting edge work and that it has great potential for learning when used in combination with chart review, particularly in teams. There is a risk of measure fatigue and a need for tools that help providers to structure their own feedback. In the future, with increased precision, the tool may have some value as a performance or quality measure, however, that possibility is a long way off. Focusing on the tools value as a learning tool in the near term was the consensus. Patients may want to know the certainty of their diagnosis, however knowing this may reduce their confidence in their providers skill and knowledge. TEP participants would like to receive study results on how ED providers feel about the tool. Working to increase the precision of the tool was identified as a priority.

Suggestions and consideration for improving the measure

1. Differentiate between hospital acquired pneumonias and community acquired pneumonias
2. Chest imaging over time
3. Time interval between the ED diagnosis and discharge?
4. Length of stay
5. 30-day outcome is not as important to them [ED providers], because it's much more removed from what they're doing. I think it's still important, but if you could look at closer outcomes, like **seven-day mortality**, that's something that I think a lot of ED studies use.
6. Perform additional chart review

**TEP participant quotes**

Perception of Sankey diagram

“I personally liked your pink and blue bar. I think that that visually helps because they're wider or less wide. And then the bottom one, negative, was very wide. So, I thought that was really helpful visually.”

“It is intuitive.”

Exploring provider reception

“In this feedback from the study is there a way to tease out how people feel about the tool? If they were just given it on their own to review without being observed? Does that take into account that aspect?”

“It would be very interesting to see how this metric would be received in a value-based care setting. In a value based care setting, often providers will routinely be getting quality scorecards and is there a risk of quality measure fatigue, or differences in the reception to a metric, a new metric, co in this case, diagnostic discordant metric, which again, providers aren't really used to. This isn't a typical kind of quality metric, so I just think that it would be interesting to expand to outside of the U of U where we're very fee for service focused, and the VA.”

“So we actually do a lot of quality scorecards in the VA. In fact, I think way too many, but I don't think they focus on these things, which is quite specific. A pushback that we get when we talk about feedback with clinicians is the time to do these activities. So we're working with AHRQ on a tool that gives structured feedback well, that you can use to sort of do structured feedback on your own cases, you can select the cases, so you could select discrepancy cases too. And we've started to talk to clinicians. And time has been the challenge. Now some of this is covid related, because I think everybody's sort of fatigued and nobody has time. But I think that may be one of the bigger challenges that we may be facing with feedback in general. I think everybody wants to improve, but where is the time to do it?”

**Finding value in the discordance measure**

1. Value as a quality measure

“For me as an attending of trainees, I think it's another one of those. Just like performance improvement, goals that we should all aim to strive for. And so as trainees are learning about what it's really like to be out there, this is one of those things, as Maya was mentioning, that could be associated with some kind of value measure or payment or something like that. So, I do think it's really important for trainees to understand that there are measures like this.”

“What I'm intrigued about is the idea of diagnostic discordance as an area for quality improvement within the framework of a logic model. How do we account for such a shift in quality in assessing quality? And are there other diagnostic discordant models out there that are commonly used for quality improvement, that maybe have a benchmark that's easier to understand, like, right now, what I'm getting is up with pneumonia, there isn't really until it's really, we're comparing against peer groups, there really isn't this a gold standard, or not a gold standard, but there's an 70%, right. But you're having to think if you get to a certain percent, then you're doing okay, or mammogram screening 100%, you can kind of wrap your head around that. Of course, it's a process measure and not an outcome measure. And for diabetes management, you're thinking a win, see, you're trying to get to below seven, you know, that's a very concrete target. But these diagnostic discordance measures, at least, especially for pneumonia, where from what you've shown so far, it's just really hard to grasp. Where one is going other than continuous improvement, which is in from a logic model perspective, does take a mental shift for a provider to say, Okay, I'm just aiming for continuous improvement, not an objective. And so that's kind of my comment, and then question on, if there are other discordant diagnostic discordant measures that have been established within healthcare, they're more commonly used perhaps, as a framework?

I'm really intrigued by the concept of diagnostic discordance as a continuous improvement measure not only in pneumonia, but across diseases, because … there aren't very many good studies that have that measure and provide providers with that information. Like, on average, to know that 60% of the time, I didn't know that. You know, you get your own sort of internal sense that medicine is uncertain and sometimes when you think that you're so certain on a diagnosis and we know that there are biases that are associated with the thought process that providers have when it comes to diagnosis. And so, I think in general there's a need for a paradigm shift in the way that providers receive information about diagnosis and diagnostic discordance. And that as a continuous improvement approach. My question would be, you know, as the body of literature continues to expand, how do we incorporate it within our quality improvement framework, because … you can't just give this to providers and expect them to all of a sudden start learning from it . It needs to be systematized in a way that is within a continuous improvement learning framework that the health system either already has, or transforms to adopt. For example, in clinical decision support, so many of you already mentioned, we use the HRQ framework for clinical decision support where it's just very linear, when it comes to translating evidence based practice and guidelines into a clinical decision support mechanism that you see an Epic, that often will even link and have a reference to which quality manager or which guideline is the foundation for that clinical decision support. And so, you know, the HRQ definitely is a well-recognized mechanism or framework for quality improvement on the department level, you know those department processes for peer review, and in the continuous improvement but not too variable across departments... and within value based care models, again, there are mechanisms for whether it's quality scorecards, but anyway, it's continuous improvement, and even rewards and potentially punishments. But rewards mostly to encourage improved quality. There are definitely rewards for health systems when it's in a value-based mechanism and if you have shared savings, you get money … if you've achieved that threshold, then the health system gets recognized and rewarded for that financially. I think that's the other part is, how do you create a system around diagnostic discordant measures as a whole? And then specifically for pneumonia as this would be a trailblazer.

I think once the AHRQ adopted this as a quality measure, and then it comes with [the] option for health systems to use it as part of their value based care or quality measure program. And the reason why we have quality measure programs most definitely is to improve quality care. Absolutely patient number one. But it's also funded by the downstream effected improved quality, which in value -ased care is very well defined on how that return on investment can be both financially beneficial, as well as quality of care beneficial... look at our counterfactual assessment that shows that there's a 3% reduction in mortality or there could be a 3% reduction in mortality if diagnostic discordance were eliminated. And that could translate into dollars. Right? So, I think, again, we’re being premature right now, but over time, I could see how it could mature.”

1. Concerns about the tool becoming a performance measure

" This is not a performance measure. It should never, never be used as one. This is only an indicator for someone to go and look to see if they could do better. Even though I know Moore Foundation wants it to become a performance measure. It's just not there. You can do whatever you want in the next six months to make it one, but it's not going to be there.”

It's not [a performance measure] right now, and it shouldn't be one unless you prove to me that the predictive value of this measure is really high, that I'm going to get a really deep signal. When I go and start looking at my charts. I don't think we've made a case that this is going to be a performance measure. Performance measure is basically the one that you'll take for accountability. That's why you want to be careful with the words that you choose here. If you're saying well, can it be a measure for improvement? Sure. Can it be a measure for learning? Sure, but can it be a measure for gauging me on my performance for accountability, pay for performance, public reporting? All of those things? No, I don't think it's there yet. It might get there with some work but it's not there yet. Cause you just have to convince me this is a very strong metric.

The science for accountability, isn't there yet to develop a kind of a measure that the clinicians are going to hate. Guess what? There's a huge pushback, and, you know, opportunity there is there. but for me, this is all improvement work.”

“I agree that right now it's not there, but I do like the direction of diagnostic discordance becoming, over time, a quality measure. And of course, it would have to go through the same scientific rigor and evidence base, as current quality measures so if you actually have a map path that out in the logic model all the way to the society. So you know, each of the quality measures has a steward and measure steward. And that's far bigger than one research group, right? It's like, it's the pulmonary society, it's the endocrinologists... So there's a long pathway to actually becoming a quality measure. And a lot of peer review and evidence based behind a quality measure… but I definitely like the idea of going beyond what our current quality measures are, which are very process oriented, and in some cases, they are disease management oriented, but they don't make us better clinicians from a diagnostic perspective. I'm not aware of a single diagnostic discordance measure. So I think that there's some value there in the meantime. So, before it gets to that level of maturity, which is super mature before it becomes a quality measure. In the meantime, as far as its weariness of providers, I mean, providers are, unless there is some psychological safety associated with any type of quality improvement. Providers are wary about performance measures, period. So it has to be in a psychologically safe environment and unfortunately, we haven't done a very good job in quality improvement in general ... in the learning case M&M would do a great job in creating that psychological safety. And there’s all sorts of structure behind it for the students and the residents that are doing quality improvement. It varies probably by the culture of the individual department and program, but presumably, there's some psychological safety there. And in value-based care and more mature measures, I don't think we've done a very good job of addressing that.”

1. Value as a signal to do something differently

“I'm not aware of any other discordance measure that is being used. And I think you've done a fantastic job of going into the deeper analysis of the meaning of this measure. What I'm looking for, what's the signal for something different to be done? What's the signal for me that I actually either missed the opportunity to do something better? Or maybe I did really excellent. Or did I over-diagnose any pneumonia? In which case, I may have given antibiotics when I should not have. So when you do your chart review, what signal strength are you finding? Even though you certainly may have only done 20 or 40 right now? What sort of the signal strength of the measure? That's what's going to help drive how many people want to use it and how many people want to find, how many people find it meaningful.”

1. Value as a learning tool for trainees & teams

“I think in general trainees, like residents, are really interested in data points as far as their performance. And they're really motivated to do well. So I think it would be a great tool for trainees. If we gave access here to trainees so they could see how they're doing as far as their discordance… but I also think that inherently both attendings and trainees are motivated to improve, which is a lot of why I think this data is really interesting in the beginning, which you learned from the providers that you interviewed. For example, in our residency our trainees really want to know how many patients per hour they're seeing, that's one of the data points that we collect for them. So this would be another thing like how much discordance are you having in your diagnose these from the ED? I think that would be a really great thing for them to learn from, and especially in the way that the tool is set up. So they could click into the chart and see what actually that patient was. And it's actually just one more thing -- the greater GME community does require trainees at least emergency medicine to do a certain number of follow up charts anyway and that is tracked by the GME folks. So that might be something for other residences as well. So this is a way that the tool can help them do their own required follow ups. But in a way that's really helpful for feedback for a patient outcomes.”

“From a systems based perspective, if there is time, and typically, it would be under a medical director guidance … that a team would look together at their quality measures, not calling out any individual but say, “As a clinic, we could do better at this particular measure by improving our process and taking those learning opportunities and operationalizing it,” but typically, if there is a financial incentive, then there's carved-out time to do that.

You could do a small group and say, “Okay, here's the case,” and then we could talk through all of the different learning points, you know, “Should we have given an antibiotic? Should we not have? What's happening here? What is the information that we had?” We do a lot of case-based discussions in small groups, every didactic so it really would be useful to have cases that we can pull from.”

1. Increased value if signal strength were higher

“I would say, you know, if you're seeing one in five, one in 10 [cases with missed opportunity], you really need to go back and use these reviews that you're doing right now, to try to figure it out, if you were going to create an actual measure out of this, how would you change it to try to get the signal strengths higher? -- how do you read out all that noise, so that you can just get that one on one and five, -- one in ten still will not be very high, but if you can get to one in three, then you can tell me for every 10 records that I look at, at least three would be meaningful, and I’ld learned something by reviewing those 10 records. So 10 really is low as a measure. I mean, still it's an indicator, the predictive values is never very high for some of these things.”

1. Increased value if learning opportunities were categorized

“I think if you can categorize what types of learning opportunities, you mean, whether it's a management issue, or diagnostic issue or documentation issue, maybe you'll come up with others. And if the signal strength for those are high, then you can it's not just the diagnostic measure, which is okay. And I think I think, to Moore Foundation, you'll say: Well, we found it as an indicator for improvement on many directions, not just diagnosis. And that's okay.”

**Patient Perspectives**

1. If their diagnosis changes or is uncertain, patients may believe that providers are not doing important diagnostic tests

I was diagnosed with hypersensitivity pneumonitis, when I was 22, went back to the same doctor when I was 37-38 I couldn't get him to do an x ray on me and then he came into the room and said, you probably should have had some follow up from the doctor who initially diagnosed you with your hypersensitivity pneumonitis, which was him, but he didn't realize at that time. So, I don't think that's wrong. I struggled to get doctors to do something on the diagnosis, I couldn't get a chest X ray. I mean, I went to every doctor that I could think of to find out what was wrong with me. And that was a three year process before I was diagnosed with pulmonary fibrosis. And whether or not they missed them, I just think that they I just couldn't get certain things done as a patient that I thought maybe would help diagnose my problem.

1. When a doctor does not give a ‘correct’ diagnosis, patients may believe that their doctor lacks knowledge or skills that another doctor may have

“Well, that would cause some concern, I understand that it's a challenge? I mean, I realized that when I was trying to be diagnosed with what I had, just because it took so long, I don't expect them to be able to diagnose everything. I just finally figured that I had something that was never going to be diagnosed until they did an x ray when they say they did the X ray. I was diagnosed from that, but I also I'm going to the university down there with the doctors, I see down Dr. Cahill, if she doesn't understand something, or she's very honest and tells me, I don't know what I'm looking at, or I don't understand this, I've got to get somebody else on board to figure out what's wrong or, and so it's not so much she's admitting fault, but she's admitting it. Maybe there's something about my diagnosis or something I've had that she doesn't understand. So she's not afraid to step outside the bounds and go to somebody else and say, I need help with this or help me to understand that, which is encouraging to me.

1. Patients may believe they have a ‘right’ to know how certain their provider is about a diagnosis

“I think honesty in the from the doctor in the diagnosis process is key to just about everything. And I felt like ever since I was down at university, honesty's been on the table from the get go... something on my blood work could show up and Barbara wouldn’t understand what it was, but she brought somebody else to find out. And she would tell me, I don't know what this is, but I know who to talk to so I can find out what this is. So I understand there's complexity there. That's, that's never been a problem. And I expect this honesty would be I mean, if you don't know, don't be afraid to say that you don't know, I'd rather hear that than say have a misdiagnosis, or you're just shooting in the dark and hopefully, it is right.”

**Suggestions for improving the measure**

Continue with chart review to identify learning opportunities and improve precision

“There's three or four things you need to be able to make sure that you collect at the end, because that's what really counts... It's all signal related. Was there a missed opportunity? Was it a missed opportunity in pneumonia diagnosis? Was it over-diagnosed pneumonia, and was it anything else in the management or any other category. Your initial 10 to 20, chart reviews always give a lot of useful information to make your chart better, and you may have to do more, because you're kind of going into a new area would be really, really useful. You've got the general gist, you were collecting missed opportunities, for the next thing, I would just focus on the last-- I realize you don't have the points. So you don't need to do the points here.

Your chart reviews are going to be illuminating, because I think the answer as to how to make them better might lie in the chart reviews you're going to do the next few months**. I think you should do more than what you're intending**, because that's where the answer is. Because you might be able to refine your measure saying, Oh, if we exclude these types of patients, include only these types of patients, we might be able to increase the signal of our measure for improvement. And once you start increasing the signal for improvement, then you're on your way to accountability.

I don't think anybody's working on a discordance measures, so you should push this hard. There're these couple of papers the Wolf and the guy from MedStar wrote, but there's not a lot of work going on discordance apart from what you're doing. So, you know, just you need to define the science basically.”

**TEP 3 Agenda**

2:00 – 2:07 **Welcome & Reintroductions**

2:07 – 2:10 **Summary of last meeting and plan for today**

2:10 – 2:40 **Presentation - Update of Progress**

2:40 – 2:45 **Precision and Recall (Slide 6)**

1. What are the implications of these values?

2:45 – 2:52 **Sanky Diagram (Slide 9 & 14)**

1. How could this information impact diagnostic learning?
2. How might we present the measure to reduce the probability of negative effects and to increase positive effects?

2:52 – 3:00 **External forces effect PPV and Sensitivity (Slide 10 & 15)**

1. Is this new information?
2. How might monitoring changing trends in the measure influence clinical practice?

3:00 – 3:05 **Associations with mortality (slide 19)**

1. Uses for the identification of correlation of false negative pneumonia diagnosis and 30-day mortality

3:05 – 3:15 **Chart review process (Slide 21)**

1. Comments on chart-review process

3:15 – 3:25 **Provider concerns (Slide 26)**

1. How might provider concerns be addressed?

3:25 – 3:35 **Implications of the measure on training**

1. As an attending who trains residents, what could you do with this diagnostic measure?
2. How could this measure be used in individual vs team-based learning?

3:35 – 3:45 **Implications of the measure on patient perception**

1. Do you foresee any unintended consequences if this measure was made available to the public?
2. As a patient, what would you do with this information, if you learned that your ED doctors diagnostic accuracy was around 50%
3. Is there a way this information could be presented to improve patient communication with health care providers?

3:45 – 3:55 **Implications of the measure on administration**

1. Existing outcome measures that may be correlated
2. As an administrator what would you do with this information if it was available for your providers
3. How might this measure impact other important quality measures both positively and negatively? (i.e. timely communication, sepsis, antibiotic overuse, patient communication…)
4. Meaningfulness of correlation, Causal inference & Confounding issues

3:55 – 4:00pm **Wrap-up**

1. RedCap questionnaire instructions
2. Next Meeting

**Measuring and Improving Diagnostic Excellence in Pneumonia**

Technical Expert Panel Session 4

* Date: September 26, 2022
* Presenter and PI: Barbara Jones
* Facilitator: Teresa Taft
* Organizer: Bahar Alimadadi
* Report prepared by: Teresa Taft

Technical Expert Panel Attendees

* Nathan Dean, Chief, Pulmonary & Critical Care, Intermountain Healthcare
* Megan Fix, Emergency Department Physician, University of Utah
* Maia Hightower, Physician Internist, Chief Digital and Technology Officer, University of Chicago
* Lacee Sims, Patient, University of Utah
* Hardeep Singh, Chief, Health Policy & Quality Informatics, Baylor College of Medicine and Houston VAMC
* Heath Young, Patient, Blackfoot Idaho, (Joined by phone, was unable to share)

Study team in attendance

* Barbara Jones
* Bahar Alimadadi
* Jorie Butler
* Alec Chapman
* McKenna Nevers
* Kelly Peterson
* Elizabeth Rutter
* Peter Tabor
* Teresa Taft
* David Classen

Introduction

The 4th Technical Expert Panel (TEP) Discussion was held virtually on September 26, 2022 via Microsoft Teams video conferencing platform. Seven TEP members as listed above were present in addition to the principal investigator, Dr. Barbara Jones and eight members of her research team. All TEP members were highly engaged and volunteered a great many insights and much advice on the project, with the exception of one member who joined by phone and found it difficult to comment. The TEP agenda is included as Appendix B.

In this report, the Technical Expert Panel (TEP) Discussion Summary is presented first followed by categorized quotations of TEP member comments, then a brief description of the presentation that was given to the TEP members at the meeting, prior to the discussion. In addition, Appendix A contains selected study results that were shared during the presentation and Appendix B contains the TEP agenda which was emailed to TEP members approximately 1 week prior to the meeting.

TEP Summary

TEP members described benefits of participating in the technical expert panel including gaining insight from hearing the perspectives of the diverse panel members, which included patients, practicing physicians, healthcare administration, and data and health informatics experts, involvement in cutting edge research and the multi-faceted user centered design process, and learning to ask better questions regarding their own healthcare. TEP members were interested in the apparent associations between study participant’s growth mindset and attitudes toward failure and their openness to the receiving the diagnostic feedback.

TEP members saw high value in implementing diagnostic measures as a tool to support providers in following up on patient cases and improving their own diagnostic skill. There was consensus that the tool is not ready to be used as a quality measure, however, with continued work and refinement a the tool may very well become a quality measure as there was agreement that the development methods were appropriate for this pathway. ED providers emphasized that appropriate treatment and disposition are drivers in the ED. There was a consensus that care should be taken that the diagnostic feedback tool is not be used to punish ED providers for uncertainty in their ED differential diagnosis.

Suggestions for measure refinement included redefining the way pneumonia is measured to incorporate positive chest x-ray and/or antibiotic treatment. Evaluating the NLP tool on a 2019, pre-COVID data set to exclude contamination with patient’s who don’t get antibiotic treatment for pneumonia diagnosis may be of value.

Panel members advised focusing future work rather than pursuing all the many directions in which this very promising project might be taken at once. Suggested focuses include: 1) refinement and implementation of the tool for provider quality improvement in pneumonia diagnosis in the ED, 2) designing a tool that uses the measure to support shared decision making discussions with patients about diagnostic uncertainty and follow-up processes, 3) adapting the tool for use in additional diagnosis types beyond pneumonia, 4) developing the tool to provide diagnostic feedback to additional hospital departments beyond the ED, 5) extending use of the tool to other institutions, 6) refinement of the measure for CMS quality reporting.

Patients want to know that they are heard and that their insights are considered when providers are making their diagnosis. They are interested in tools that help physicians to rule out inaccurate diagnosis and to focus on a patient’s most urgent health problem. Patients may appreciate knowing that their ED provider followed up to see how they fared after leaving the ED.

Suggested implementation strategies include engaging a clinical champion who works in the unit who can promote use of the tool and developing a security plan and a sustainability plan to ensure that the tool remains funded and up-to-date. A barrier to use would be if providers find the tool cumbersome or inefficient to use, therefore making access efficient is important prior to implementation. A possible unintended outcome is the measure may drive changes in patient care or documentation as providers seek to improve their ratings.

With the development of automated diagnostic clinical decision support (CDS) tools, the way that ED diagnosis are made is changing. Outside research is being done to measure effects of automated diagnostic support tools that might inform or be incorporated into the project. Sharing diagnostic certainty with patients is being explored elsewhere. Intermountain Healthcare’s 7-day hospital readmission measure for assessing outcomes of ED discharged patients is an example of how the measure might be extended beyond hospital admission and discharge. AHRQ is releasing a tool that helps providers identify interesting cases for review.

Quotations

This section contains quotation the TEP members shared regarding the following concepts: 1) Experience of participating in the technical expert panel, 2) Impressions of study results regarding growth mindset and participant comfort with failure in the context of provider openness to diagnostic feedback, 3) Value of the feedback tool for diagnostic quality improvement versus as a quality measure, 4) Suggestions for Tool Refinement,

5) Implementation strategies, perceived barriers to uptake, and possible unintended consequences, 6) Next Steps Planning, 7) Patient priorities regarding the future direction of the project, and 8) Informative outside research. Illustrative TEP member quotes are included after each of the seven summaries.

1. **Experience of participating in the technical expert panel**

*DC: “****It’s been a wonderful project for me because this, to me, is the future****, which is all automated measures of quality and safety. CMS has basically said by 2030 will all be reporting only automated measures, so I've really enjoyed this, and all the different perspectives, and I think the measure that we've developed here has a good likelihood of turning into something that might ultimately be a national measure. So, for me****, it's been quite exciting****.”*

*HS: “****I always enjoy hearing what practicing physicians have to say around some of these topics*** *of measurements and diagnostic discordance and learn a lot from their real-world insights.”*

*ND: “I've been very interested and think* ***I've learned much more than I have contributed.*** *The whole issue of diagnosis in pneumonia has been understudied and difficult.”*

*“****The thing that's been most fascinating has been the multi-user design focus and I think that definitely is the wave of the future, present and futur****e, and how we create experiences or design experiences around the patient experience, the consumer experience, the staff and faculty care team experience and really being able to touch upon all of these different users with a similar intervention, taking it to those diverse perspectives and coming up with a unifying intervention. So, I think that's fascinating and I look forward to more multi-user design centered interventions.”*

*MF: “It's been such an honor to be a part of this group****. I have loved hearing all of your different perspectives.*** *It's definitely been interdisciplinary, and I've loved having the patient’s perspectives on the panel as well. And I think one of the things that's been great for me to see, just being involved in a very robust project like this, Is just the many different layers and the multimodality approach going from the you know the data aspect, and then having the panel, and then having the physician interviews; and I've learned a lot about a concept that I knew before, but I never really put a name to it, diagnostic concordance and discordance… so,* ***it's been really interesting****. I appreciate all of your hard work because I've really been impressed with all of the different avenues of how many people are working on this project so hard. So,* ***it's been really impressive****.”*

*LS: “I took a little trip to the emergency department not too long ago and had pneumonia, and I think I asked better questions, just to the doctors. Because I've had pneumonia so many times, but it was a little different this time because it was COVID pneumonia, but* ***I think that I was able to ask different questions in a different way, about my treatment and ca****re and stuff like that; and I believe in continuing education and you guys keep saving me and I say, “keep doing a good job,” because I appreciate it”.*

**2. Impressions of exploratory study results regarding growth mindset and participant comfort with failure in the context of provider openness to diagnostic feedback**

*ND: “I was very interested in that psychological profile and the differences in the results. In my own work it’s been suggested that the use or lack thereof be stratified by psychological profiles, and I think that there would be an association. I don't know whether it be MMPI or something more sophisticated like this, but I think that's fascinating.”*

*MF: “I was really interested in the differences between the VA and the UU population, the difference in documentation, differences in the amount of false positives, and usually with U of U having a lot more. And I was thinking, you know, everybody's an individual and it's kind of different in terms of the culture where you practice… it all comes down to the individual… the analysis by growth mindset and failure acceptance-- that was fascinating because that is an underlying, I think, understudied component of change that could be applicable to a wide variety of initiatives when it comes to the sense of usability versus whether you're growth mindset or not... It’s potentially really beneficial for people's own internal quality improvement, but then … it came back to the individual, right? Are you a growth mindset person who's really going to want this or are you not?”*

**3. Value of the tool for diagnostic quality improvement versus as a quality measure**

*MF: “When you're in the ED, you have limited information and so you're thinking that maybe three diagnosis are most likely on your differential, and in those in the highest possibilities you may not have the definitive diagnosis. And so, what they're saying is that treatment, and disposition, and making sure that you're covering your bases is more important at the time; then the eventual discharge diagnosis may be days or weeks later.”*

*ND: “it's quite clear that I don't often know what's wrong with the person who's injured, maybe, a part of their body, but we have to make disposition and treatment decisions for the various possibilities that could be there.”*

*ND: “It kind of shocks me that apparently your docs ranked [diagnostic accuracy in the ED] after treatment decisions, because you can't get treatment, right if you don't have the right diagnosis.”*

*DC: “it seems to me there was a great deal of interest in using this as an internal quality improvement tool, even though it was a flawed measure…. and what that tells me is that this has really not been an effective quality improvement area for physicians, right? They never get any feedback, right, and they would love some. So, you know, I think that's a real aha here, that even with an imperfect measure, people still are interested in it and where it leads me to go is not only could it be used as a quality improvement measure, but what if you made it real time and could give physicians information like this in real time?”*

*HS: “the signal strength: out of every five charts I look at, one's going to have something really useful [for a diagnostic learning opportunity]; … do I have the time to look at those five charts is really the question… It was not surprising to me at all… when you start looking at records, you'll find that a lot of this is just noise. So, it's 80% noise, roughly 70 to 80% noise”*

*ND: “I'm just going to say, an imperfect measure, if it's not threatening to physicians as being potentially a punitive or comparative thing, they're willing to put up with this because they want to learn and are interested in it; but as soon as it becomes a performance standard or measure, everybody gets damn nervous about it and there've been too many examples of negative consequences of performance measures…”*

*MF: “I'm thinking about the whole idea of using it more as an internal quality measure and not as something that's punitive because, there're so many variables and we're talking about having the chest X-ray at the time of Ed. So, often the chest X-ray says there could be an infiltrate: clinical correlation advised; and you just don't know yet, in the Ed, if it's going to blossom into a pneumonia or not. And so, we are highly encouraged to cover, empirically...”*

*HS: “It will not stand for a quality measure. You won't be able to convince NQF or anybody about this. So, you either have to refine it to get the PPV higher, make your validity a stand out. But as an internal quality improvement tool, you know it should be, you know, maybe close to where it is.”*

*MF: “based on very unclear chest x-rays, often you know in a patient who's very sick. And I know you talked about people who had changes in their disposition from Ward to ICU and etcetera. But I just think that there are so many variables that are at play when you make that initial diagnosis of possible pneumonia. And as you mentioned, there're lots of big lists of all these things, but we are much more scrutinized for not giving antibiotics in the setting of a possible pneumonia that may develop over time… Then the other way around. So, I just think it would be much more interesting for providers to say, oh, right, I remember this case. And yes, what ended up happening? Well, they ended up having a PE or whatnot, right. So I do think we have to be careful about the message.”*

**4. Suggestions for tool refinement and validation**

*DC: So, as you know, in our research work, we've always defined pneumonia in the emergency department as having to have a combination of clinical diagnosis and abnormal radiograph. But here you've got just a clinical diagnosis in the ER, which leads to the problem of hospital acquired ammonia, VAP prior to discharge as well. Likely less accurate diagnosis and emergency department. [BJ: Pneumonia diagnosis could require a chest X-ray, but then you would actually lose the ability to measure this as its own quality measure.]*

*HS: What was the reason why only 2/3 of patients got an antibiotic when they were supposedly diagnosed with pneumonia. That kind of calls into question the methodology a little bit of the data. [BJ: One other way you could refine this measure is to actually require that they receive treatment. I will remind you that a year and a half of our data were during COVID and for COVID pneumonia, you don't actually get any of that.*

*… It is important to note that yes, a lot of patients who receive a label of pneumonia in the emergency department might not actually be getting treatment for pneumonia.]*

*[BJ: “we could keep going down the measure development pipeline for quality… the chest image confirmation of a positive diagnosis to me seems the most compelling and kind of accurate, useful] HS: within the first maybe like 48 hours or so? [BJ: and possibly even also against the discharge diagnosis. They really ought to have a positive chest imaging too. So, the you know that to me seems like the best candidate measure for like an actual quality measure.] HS: Wait, so you mean to say discharge diagnosis date? Like, let's say they got discharge 10 days later, you look for a chest X-ray with pneumonia like 2 days prior or something? [BJ: So we also talked about the false negatives, which are a very interesting group, you know, but some of those are hospital acquired pneumonia, right? So you could refine that by taking the chest imaging and requiring that to be positive in the first 48 hours, but then actually thinking about, ‘What is a really outstanding quality measure?’ … There are actually quite a few cases where the initial diagnosis is positive, the discharge diagnosis is positive and the patients have no chest radiograph confirmation of pneumonia at all, and to me that's like a system problem. It can happen, but it probably shouldn't happen very frequently. So, I thought that could be a pretty good candidate. A lot of them are very short admits that really just have clinical inertia and they [discharging physicians] just don't change their diagnosis even if they actually are found to have no evidence of pneumonia on chest imaging. So… in this, you know, 17% of all discharge diagnosis lack a confirmatory X-ray.] HS: So, you know all the uncertainty stuff came up, right. And the interviewees, they said, “Look, a lot of times there's uncertainty and the diagnosis also will evolve overtime. So, if you add the chest X-ray to the Ed-- so a clinical diagnosis plus chest X-ray and see if that's very different and maybe you can test that in the dataset that you already have. You just have to change the parameters and make sure that you're adding like the within 48-hour X-ray. So, I'm just thinking you can probably do some refinements. I would try to exclude COVID though because it looked like some COVID patients. I know I heard something around. I was like, oh, this is not getting messed up. A lot of our studies we did before covered even though they were funded after COVID we said because it'd be so much of contamination from COVID that we won't be able to account for it. And so, I don't know whether you can do it on a 2019 dataset and see if it works better. [BJ: We definitely could. You could also bring in virology and micro information if you were really wanted to distinguish a viral from bacterial.”]*

*ND: “I think though, the whole bugaboo about diagnosis is “What's the gold standard?” And so, this remains, probably, the most challenging area of pneumonia work research.”*

*ND: “In terms of improving the performance measure, I think combining the Ed radiology findings with the clinical diagnosis, one way is looking at antibiotic use. And with the problem of COVID, maybe looking for Remdesivir, steroids, immune modulating drugs typical of COVID might be one way to do that. And certainly, also, Tamivir prescribing I think might help that considerably. Because if you've got leftover patients with pneumonia that you're not treating and also if you excluded the Hospice Care patients, then it really should be a pretty small number that are leftover.*

**5. Implementation strategies, perceived barriers to uptake, and possible unintended consequences**

*MH: “Implementation of the tool in the ED would require tested usability, a clinical champion who will promote the tool with colleagues in the ED, technical security, and a sustainability plan for ongoing financial support and staying relevant… having a champion, somebody in the emergency department … who finds that it's of value and can really share with their colleagues, especially as a learning tool, how valuable it is; I think that's number one. Without having that champion, or that that advocate, it's going to be really challenging … And two is the technical infrastructure making sure that we have a tool that has high usability, and that already has been shown, that the providers really like the look and feel of the tool and it provides the necessary information. So, I think that hurdles already been surpassed. And then, the typical kind of technical security and all the other technical components. And then from a sustainability you know, “What is this product?” “Is it going to be continued as a research project product or is it something that is supported?” “Who supports ongoing development of the tool?”*

*MH: “something that allows it to scale beyond one institution. I think all those are really helpful as well because then you end up with a tool that starts off very nicely. If we don't have that infrastructure or support ongoing, it starts off really nicely, but through just technology evolves, and what was once a diamond becomes dust. So for anything that's that's, you know, homegrown or innovation that starts off as an innovation project. So I think those are some of the, the things that are and then of course further enhancements. So that's where if pneumonia was the 1st. And then there's other diseases that Ed physicians or family medicine physicians or internists want to learn from. Then I think it it continues to grow as a tool.”*

*MF: “I think just resistance on the part of the providers and lack of time. I mean you spend a lot of time setting up these interviews with the providers and going through everything with them. I think if we release this to a group, say we have 50 people in our group, I think it's going to be really hard to get people to do it without a champion, or someone else, or sitting down with Barb. Something like that. I just think people are very busy and there's resistance to doing a lot of extra things.”*

*MF: “the unintended consequence could be that people are really scrutinizing their cases and not providing antibiotics when we may not know. So, you may find that providers practice changes, patterns change based on what they've been seeing or learning from the tool, and maybe that is what we want. I don't know, but we don't know what that will cost.”*

*MH: “unintended consequence could be a change in behavior… where if there's diagnostic uncertainty already inherent and the diagnosis of pneumonia, that providers start narrowing their creativity or their approach to pneumonia to a standard of care that hasn't yet been proven to be the best standard or the evidence based optimal standard of care. So, I think that's the part that's challenging, that a tool, especially if it's within workflow, may truly change physician behavior and that's a national concern on how much should clinical decision support tools, especially when AIML tools could be literally clinical decision support tools versus automation. And this is definitely one where it should be decision support, but I don't think our clinicians always have that same level of nuance, an understanding of the difference between clinical decision support and the computer told me to do it.”*

**6. Patient priorities regarding the future direction of the project**

*LS: “For us, for professional patients, listen to what we have to say. We've kind of been through it all. We've got a list of doctors and a list of diseases … listen to what they're saying, even if it's something small, it might be something significant that you know, that they just kind of brushed over, but it might be part of the bigger picture… just listen to the people what they're saying because sometimes we have a little bit of insight too.”*

*LS: “I think they [providers] should start with the issue in front of them and listen to how the patient feels, even if something seems small or insignificant, and then look at the medical record.”*

*LS: “I am going to keep the goal of creating measures that encourage providers to listen!”*

*LS: “I was really happy that he just said, look, I know you've got all this, but we're just going to deal with the one problem. And so, I don't know if the if the tool was there it you know, if he's like OK look, you know, I can go through ABC and D and see what we can rule out right off the bat and then go from there.”*

*LS: “The patients that are upstairs, did the ED doctors check on them? I know that [Presenter] had mentioned, sometimes they didn't have access to the that information anymore. I mean, I wonder, do we ever cross your mind when we go upstairs?”*

**7. Next Steps Planning**

*DC: “I think there's so much work to do either as a quality improvement or a quality measure, I wouldn't distribute the work into broader areas of disease, right? You’ve got enough work to do here and I worry if you go into a million other diseases, it'll be, you know, a mile wide and an inch deep. So, I'd actually push back very hard and say focus here, you’ve get a lot of work to do.*

*DC: “stay on the pneumonia. If you jump into epidural abscess or other things, I think you will diffuse pretty quickly.”*

*MH: identify which user stories you really want to capture and encapsulate and the next iteration of the project, collect more of those voices at the table… whereas the Ed docs are maybe saying, you know, “I don't want a quality measure, I want a quality improvement tool where I am able to self-assess my own improvement over time,” and maybe it's benchmarked, maybe it's in real time, we don't know. We would need to understand that the Ed physicians or physicians in general if we were to go to other specialties what is it that would be useful? What pain point does this tool solve? And then for patients … about pain points that she would like solved with this tool as well. And so I think that's where you can go in a lot of different directions, but it all depends on which user story is going to be more dominant in the direction you choose to go… usually when we look at these research studies, we're looking from the lens of, say our faculty research academic lens. But it may be helpful … even this group is skewed in that direction, in really identifying which user stories you really want to capture and encapsulate and the next iteration of the project, collect more of those voices at the table.”*

*MF: it's really important for us to make the diagnosis as best as we can. We're discharging them. Because, as you mentioned, they're not going to have continuous care in the hospital. Also, the outpatients are a much less sick group, but there're so many times where you're sending people home on empiric oral antibiotics for outpatient management. And I will follow up on them routinely too, and see if they followed up with their PCP or if they didn't, who knows? But … if we could expand and get quality data on the outpatients, which I'm sure is going to be challenging because many of our patients don't follow up and I'm sure that will be difficult to get the data, but if we could, I think it would be very interesting.”*

*MH: “when it when it comes to expanding, it depends on how you would like to expand if it's using the exact same tool, then it's revalidating with other subspecialties… if it's expanding to a different disease state, you can still use the same methodology to apply to new disease states. I think the methodology itself is extremely robust and developing -- I know the intent was to develop a quality measure, but we may land nicely on a bunch of quality improvement tools for different diagnosis. But I think that it all depends on what you're interested in doing, and clearly some point parts of the qualitative versus the quantitative analysis are better scaled. Essentially doing the exact same methodology, but either for a different specialty or the exact same methodology for different disease state and that would make the tool more robust.”*

*DC: “having developed a lot of measures for CMS electronic and quality measures. This is just the pathway we went through. You develop your initial measure, it has problems, you refine it, you add inclusion and exclusion criteria, and then you move down the road. So, I do think there's a great opportunity there as well.”*

*HS: “I think there's a resource that's coming out from AHRQ … it's a tool that physicians of any specialty can use to systematically go through their own cases and use them for learning and improvement. It's not-- there's no question about bad or good. You pick the five cases that you want to look at every quarter or so and just look at the ones that you think are high risk and you might learn from them … It's like, I don't know, a 1520-page long thing. It’s probably a longer run now just before release, but it gives lots of examples of what you could do. When we pilot tested this tool, one of the biggest pushbacks we got was, “It's a great tool, but we're not going to have time to use it,” or, “Just get the residents to do it, because this is what they should be doing.” And it was like, “Oh, I don't need that feedback.” But, you know, they really do; all positions do. So, if you can implement something like this as a package, and [include] other types of cases they can look at, not just pneumonia, but other types of cases.”*

*DC: “If we just start with the issue of pneumonia, one goal could be a tool to help physicians improve the quality of care they give pneumonia. Another could be a tool that informs patients and allows them to participate in the decision-making process. So what if you created a version of this with a patient front end, that physician and patient could go through and see the uncertainty and maybe help arrange the follow-up. And then obviously a third goal of this, if you stuck with pneumonia, is some sort of quality measure. I would see all three goals as possible, but I think you can do all three with what you've started with. I guess my I would come back to my boy as long as you stay on the ammonia. If you jump into epidural Abscess or other things. I I think you sort of you will diffuse pretty quickly.”*

*DC: “first as a quality improvement tool and then as a quality measure tool. So, I that's how I’d probably frame it, as a quality improvement tool. I think we've already acknowledged that even in its current format, flawed as it is, doctors like to see it. So as an internal quality improvement tool, I think it has value as now, but I think you've outlined several ways you could make it more useful. So, I would strongly suggest that you go down that road and I don't think you have to make it too much more accurate to make it useful, so I’d really do that.”*

*MF: “I think you're right… there's a lot of places that we could use this, --not just pneumonia, right?”*

*ND: you could do this with hospitalists, you could do it with Instacare/urgent care clinic providers. Those would all be really the same thing … the issue with the Instacare/ urgent care is they have some such limited data and testing compared to the other settings. The hospitals on the other hand have an advantage over the ER in that the tests are available, and you do have follow up.*

*HS: I think we're still talking about this being sort of some type of a measure for internal improvement in order specific quality improvement measure and I think because it was tested in the Ed setting. It's probably should stay with the ED; however, it may could be combined with additional cases that would also be useful for VA physicians to look at, not just these pneumonia cases, but they might learn from some other similar types of cases. I mean, there's a ton, right? So everybody who you see who gets admitted for spinal epidural lapses a week later, you should probably look at it and see if you missed the opportunity to see them, and there're other cases.”*

*ND: “I was just going to say something additional that I think to use this as a quality improvement tool, it has to be figured out-- how to do it in a very timely way, as particularly the nuances of diagnosis of an individual patient aren't necessarily in the chart and the physician will not remember more than a couple weeks, seeing a patient and be able to do feedback. And so, it would have to be automated. If it waits for discharge, diagnosis and coding, that's often, apparently, delayed.”*

*HS: “The point about adding something to the portfolio to give more breadth for the physicians to improve because they not only need to improve in just pneumonia, they need to improve on like 25 other things as well like everybody does. So, it's not just pneumonia yeah, you can focus on the pneumonia and Barbara, I recall the conversation around well, should you stay on the same or should you go to outpatient, followed by an admission type of measure, something like we've done. In family care, diagnostic error studies, pneumonia was in the top few that at least in family care it shows up. I'm sure it shows up in ED as well… So I think you have to sort of make a decision. If you want to focus on pneumonia and you want to keep developing it like David saying, you could keep developing it, but then you could also think about expanding a different measure to try to get into some of those readmissions and primary care or ED.”*

**8. Informative Outside Research**

*ND: “we're using, basically, artificial intelligence to read the X-ray in real time… I just would ask you is that going to fundamentally change our misdiagnosis when we see the radiology interpretation real time or not? …The first step of the tool is to calculate a percent likelihood of pneumonia based on 40 demographic and clinical factors, and the radiology report for this, then the AI (artificial intelligence); and so the Ed physician has that to compare to their own clinical diagnosis.”*

*ND: “Our next AHRQ grant submitted has patient centered aspects giving them a measure of certainty in diagnosis and information about the site of care decision.”*

*ND: “A straightforward electronic measure that we've used in outcomes [at Intermountain] is 7-day secondary hospital admission for patients sent out from the Ed. And it misses the patients who got well spontaneously regardless of what their treatment was, but I think that that could be a measure that could be added to Barb's work here, in terms of diagnostic accuracy, even though crude, it is electronic.*

*ND: “we have two databases comparing clinical outcomes when the e-pneumonia clinical Decision support tool is available to edit providers versus not. And we haven't looked at diagnostic accuracy because the difficulty of doing chart review by committee for thousands of patients as we're involved in these studies is overwhelming, but it might be just interesting, as an exploratory secondary analysis, if there an automated electronic method to look and see if there is a signal because that's one thing that's come up with reviewers too. We've shown improve clinical outcomes and reviewers have questioned, and said, “Well, are your patient populations diagnosed with pneumonia different when the e-pneumonia was available versus not?””*

Presentation Summary

The presentation shared with TEP members is presented here in brief, organized as Background with the measure concept and study aims, Methods, and Results as a summary of both quantitative and qualitative findings. Selected tables shared during the presentation are included in Appendix A.

Measure Concept

In looking at electronic health record (EHR) data, the research team noticed many changes between the diagnosis of pneumonia from the initial emergency department diagnosis, when compared with two other places where pneumonia is diagnosed, the chest imaging and then also the final or discharge diagnosis. So, using Veterans Administration (VA) data (diagnosis codes) and natural language processing techniques for automation, the team developed a measure of the discordance between the initial, ED, pneumonia diagnosis versus the radiographic diagnosis or the discharge diagnosis and the denominator is all patients that are hospitalized from the ED. This measure can be summarized at the facility and the provider levels and provided to physicians as regular feedback and the tools that were developed support investigation at the individual case level. The research team applied the NLP tool to University of Utah EHR data to see how well it performed, and then customized that NLP tool to assess how much customization was needed to improve performance.

Overarching goals of this project were defined in short, medium, and long terms at the provider patient and system levels as shown in Table 1.

Table 1. Overarching goals

|  |  |  |
| --- | --- | --- |
| **Goals of a meaningful pneumonia diagnosis measure** | | |
| Short-term | Medium-term | Long-term |
| Provider  o  **Increase access** to:   * patients’ clinical course * diagnostic discordances * differences in personal and department diagnostic alignment * 🡪 Reduce fragmentation   o  Recognition of personal diagnostic patterns  System  o  Increased awareness of prevalence and associations with quality and outcomes   * Track interventions   Patient & Provider  o  Increased recognition of diagnostic uncertainty | Provider  o Enhanced diagnostic skills  o Increased self-efficacy  o Improved workup and documentation  o Increase accountability and attention  Patient  o More appropriate, timely treatments  o Reduction in antibiotic overuse  o Increased self-efficacy (to question a diagnosis)  System: Professional societies and quality organizations  o Adoption of diagnostic alignment measure | Provider  o  More accurate diagnoses  o  Reduction in burnout  Patient  o  Improved health outcomes  o  Increased satisfaction  System  o  Enhanced public trust  o  Reduction of antibiotic resistance |

Specific Aims consisted of 3 parts. Empiric Testing, Chart Review, and User Testing.

Empiric testing aims were:

1. Evaluate associations between other important clinical measures
   1. Other treatments: antibiotics, diuretics, steroids)
   2. Outcomes: ward🡪Intensive Care Unit transfers, inpatient and 30-day mortality
2. Evaluate relationships between diagnostic discordance and patient outcomes
   1. Account for multiple patient factors, that can confound relationship
   2. Contrasted the observed outcomes versus the outcome that would be expected based upon patient characteristics alone (were the patient to not have the discordance).

Chart Review aims were:

1. Assess validity of the measure
   1. “do clinicians agree with the measure?”
   2. Is the measure representing what it is intended?
2. Determine Inter-rater reliability
   1. “do clinicians agree with each other?”
3. Identify and characterized sources of diagnostic discordance
   1. Do these represent error?
   2. Do they represent learning opportunities?
4. Assessment level of uncertainty in ED providers a pneumonia diagnosis

Usability testing aims were:

1. Evaluate feasibility and usefulness of the tool through semi-structured interviews and the System Usability Scale survey
2. Elicit providers reaction to the tool
3. Explore psychosocial attributes of providers that might influence reactions to data with Growth Mindset and Failure Attribution survey questions.

Three methods for testing the tool were reported. 1) Empiric testing with Veterans Administration (VA) data (1/1/2015 – 1/1/2022). 2) Chart review by 2 clinicians of 50-100 charts from VA and UUH to identify presence of pneumonia in initial diagnosis, chest imaging results, and discharge summaries. 3) Usability Testing at UU with 11 ED providers.

Summary of findings

* Measuring diagnostic discordance is:
  + Feasible to implement across 2 healthcare systems
  + Reliable between 2 raters
  + Valid (although validity is decreased compared to individual components of measure)
  + Associated with other important clinical measures
* Diagnostic discordance in ED may not be a perfect quality measure
  + Uncertainty and competing priorities inherent in pneumonia
    - Lack of gold standard of diagnosis
    - Competing priorities with treatment/disposition
  + Different styles of documentation/use of diagnosis (VA versus UU)
  + Concerns about using it as a benchmark/performance measure
  + Few represent missed opportunities, but many represent learning opportunities
* Feedback around diagnosis may be useful to providers for learning
  + Fills a need for case review/reduce fragmentation
  + May be more useful if it includes other diagnoses
* Analysis of the ED provider interviews revealed the following themes.
* Theme 1. Diagnosing pneumonia. In the ED context is uncertain and in tension with other priorities.
* Theme 2. Existing diagnostic skill improvement processes are fractured.
* Theme 3. Clinicians liked the feedback tool and features.
* Theme 4. Clinicians had strong reactions to feedback data.
* System Usability Scale ratings
* The median SUS score was 75 (well above average) with a range of 52.5 to 92.5.
* Correlations of Feedback and Usability Ratings with Growth Mindset and Failure Attribution
* The sample group were on average somewhat comfortable with failure and somewhat growth oriented.
  + Lower comfort with performance failure was associated with :
    - Median low comfort with tool feedback
    - Low tool usability ratings
  + Higher comfort with performance failure was associated with
    - Wide range of comfort with tool feedback
    - A wide range of usability ratings, trending high
  + Growth oriented mindset was associated with:
    - A wide range of comfort with tool feedback, however more were very comfortable
    - A wide range of usability ratings, trending high
  + Low growth mindset scores were associated with:
    - Less comfortable with receiving tool feedback
    - Very low or very high usability ratings

**Appendix A. Selected Result Tables and Images**

**Table 2. Precision and recall of the VA trained NLP tool in VA data**

|  |  |  |
| --- | --- | --- |
| **Department** | **Precision** | **Recall** |
| **Emergency** | **88.1** | **86.0** |
| **Radiology** | **71.4** | **96.2** |
| **Discharge** | **88.3** | **93.0** |

**Table 3. Precision and recall of the VA trained NLP tool on UU data**

|  |  |  |
| --- | --- | --- |
| **Department** | **Precision** | **Recall** |
| **Emergency** | **84.7 (-3.4)** | **94.3 (+8.3)** |
| **Radiology** | **78.3(+6.9)** | **100.0 (+3.8)** |
| **Discharge** | **65.5(-22.7)** | **92.7 (-0.3)** |

**\*”(+/- difference)**

**Table 4. Precision and recall of the VA trained NLP tool on UU data**

|  |  |  |
| --- | --- | --- |
| **UU**  **Customized**  **(11/30/2021)** | **Precision** | **recall** |
| **Emergency** | **89.3 (+ 2.3)** | **94.3 (+8.3)** |
| **Radiology** | **87.0 (+15.6)** | **100.0 (+3.8)** |
| **Discharge** | **75.0 (-13.3)** | **95.1 (+2.1)** |

**Table 5. Treatments and outcomes associated with false negative and false positive diagnosis of pneumonia in the ED when compared with hospital discharge diagnosis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Event** | **Observed** | **Expected**  **(logistic regression)** | **Expected**  **(XG-boost estimation)** |
| **False negative compared to discharge** | | | | |
|  | **Receipt of abx within 24 hours** | **66.8%** | **91%** | **91.7%** |
|  | **Receipt of steroid** | **31.8%** | **35.3** | **34.6%** |
|  | **Receipt of diuretic** | **29.3%** | **26.8%** | **27.7%** |
|  | **Ward- ICU transfer within 72 hours** | **9%** | **6.9%** | **7.5%** |
|  | **Inpatient mortality** | **7%** | **5.4%** | **5.7%** |
|  | **30-day mortality** | **12.6%** | **10.5%** | **11.3%** |
| **False positive compared to discharge** | | | | |
|  | **Receipt of abx within 24 hours** | **67.5%** | **45.1%** | **47.2%** |
|  | **Receipt of steroid** | **31.0%** | **20.3%** | **22.8%** |
|  | **Receipt of diuretic** | **34.1%** | **32.8%** | **33.1%** |
|  | **Ward- ICU transfer within 72 hours** | **5.5%** | **6.2%** | **6.0%** |
|  | **Inpatient mortality** | **4.3%** | **3.3%** | **3.6%** |
|  | **30-day mortality** | **9.3%** | **7.9%** | **8.3%** |

**Table 6. Chart review measures of reliability and validity**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3.1.2b. Chart review validation – inter-rater reliability between 2 clinician reviews (emergency physician and pulmonary/critical care physician)** | | | | |
| **Concept** | **Prevalence** | **IRR (Kappa/ weighted kappa)** | **Se, PPV**  **– Barb** | **Se, PPV - Liz** |
| **Initial (ED) diagnosis of pneumonia** | **.62** | **0.88** | **.87, .96** | **.89, .89** |
| **Chest imaging diagnosis of pneumonia** | **.64** | **0.87** | **.91, .91** | **.90, .81** |
| **Discharge (DC) diagnosis of pneumonia** | **.60** | **0.88** | **.96, .78** | **.96, .70** |
| **FP Discordance with discharge** | **.28** | **0.78** | **.92, .92** | **.75, .75** |
| **FN discordance with discharge** | **.24** | **0.88** | **1, .52** | **.9, .58** |
| **FP discordance with chest imaging** | **.14** | **0.56** | **.8, .67** | **.4, .33** |
| **Certainty of initial (ED) diagnosis)** | **NA** | **0.85** | **NA** | |
| **Certainty of discharge (DC) diagnosis)** | **NA** | **0.87** |
| **Saferdx “missed\_opportunity\_score”>20** | **.28** | **0.31** |
| **Saferdx “missed\_opportunity\_score”>40** | **.10** | **0.30** |
| **Case represented a learning opportunity** | **.62** | **0.35** |

**Image 1. Reviewer ratings of certainty in their pneumonia diagnosis.**

**Chart

Description automatically generatedChart, bar chart

Description automatically generatedChart

Description automatically generated**

**Image 2. Growth Mindset Correlated with Feedback Comfort**

**Image 3. Failure Comfort Correlated with Feedback Comfort**

**Image 4. Growth Mindset Correlated with Tool Usability Ratings**

**Image 5. Failure Comfort Correlated with Tool Usability Ratings**

Appendix B

**TEP 4 Agenda**

2:00 – 2:03 **Welcome**

2:03 – 2:04 **Review Agenda**

2:04 – 2:10 **Reintroductions: Share Insights Gained by Participating in the TEP**

2:10 – 3:00 **Presentation – Project Progress and Study Results**

3:00 – 3:55 **Discussion**

1. What stood out for you in the results presented today?
   1. Did anything surprise you?
      1. Quantitative results
      2. Qualitative results
2. What are the strengths of the measure?
   1. What are the weaknesses?
3. In what ways might we refine the measure to make it more useful or accurate?
4. What would need to occur for us to actually implement this diagnostic feedback in the ED?
   1. What challenges should we anticipate?
   2. What are the potential unintended consequences of implementing this measure?
5. What are possible opportunities for expanding the use of the measure?
   1. Which setting would be most amenable?
   2. Setting-specific strengths and weaknesses?
6. ED providers stated that diagnostic accuracy is of secondary importance in the ED, relative to proper treatment and disposition. Please explain why you agree or disagree.
   1. What are the pros and cons regarding continuing to use and further refine this measure in Emergency Departments?
   2. What other diagnostic quality measures could also be relevant to ED providers in addition to the measure we chose?
7. What is the ‘TRUE GOAL’ in developing the measure? Where do we go from here? We have a measure diagnostic discordance. How do you get from diagnostic measures to improving quality?
8. Do you have an additional comments, suggestions, or questions?

3:55 – 4:00 **Wrap-up**

**Agenda and Methods for TEP #5**

**Measuring and Improving Diagnostic Excellence in Pneumonia:**

**Technical Expert Panel (TEP)**

**MEETING AGENDA**

**Date:** Friday, January 26th, 2024

**Time:** 10:00 am-12:00 pm

**Location:** Microsoft Teams

**Questions/support:** Bahar Alimadadi: [bahar.alimadadi@hsc.utah.edu](mailto:bahar.alimadadi@hsc.utah.edu) cell phone: 801-695-6726

10:00 am-10:15 am **Welcome and Introductions**

* + Name, expertise/experience relevant to the project
  + What impact are they hoping to have on today’s measure and/or the world of pneumonia diagnosis (and care?) in general?

10:10am-10:45am **Project Presentation**

10:45 am-11:50 am **Questions for the panel/discussion**

*Panelist experiences:*

* What are your past experiences with quality in pneumonia diagnosis and care?
* What are your past experiences with existing quality measures or quality improvement programs in pneumonia?

*Measure usability/actionability:*

* + Please review the logic model (slide 21)
  + What actions would you expect a hospital to take to improve on the measure?
  + What settings patient factors might influence how well a hospital performs on this measure?
  + How would you use the information of how a hospital performed on this measure?
  + Feasibility
  + What people/processes/structures within a healthcare system are required to use the measure as specified / “at the bare minimum”?

*Impact of the measure:*

* Please review the overarching goals (slide 22).
* What impact do you think this measure will have on our bigger goals for patient care and outcomes?
* Can you imagine any unintended consequences on these goals or others?

*Anticipated changes in patient care*

* How do you think performance scores on the measure might (or might not) change with the use of new imaging methods? What would be the expected impact on quality if these new imaging methods become more common?

11:50 am-12:00 pm **Next Steps**

**Chest Imaging-Confirmed Diagnosis of Pneumonia**

Technical Expert Panel Session 1 (Phase 2 of Measure Development)

Date: January 24, 2024

Presenter & PI: Barbara Jones, MD

Facilitator: Lindsay Visnovsky

Organizer: Bahar Alimadadi

Report prepared by: Lindsay Visnovsky, Bahar Alimadadi, and Barbara Jones

Technical Expert Panel Attendees:

* **Mark Metersky, MD, FCCP, FACP** – American Thoracic Society (ATS) member; Chief of Division of Pulmonary, Critical Care and Sleep Medicine, University of Connecticut; Expertise: pulmonary and critical care medicine
* **Melinda Neuhauser, PharmD, MPH** – Office of Antibiotic Stewardship, Division of Healthcare Quality and Promotion, Centers for Disease Control and Prevention; Expertise: antimicrobial stewardship and eCQM priorities for stewardship-related measures
* **Julio Ramirez, MD, FACP** – Infectious Diseases Society of America (IDSA) member; Chief of Division of Infectious Diseases, University of Louisville; Expertise: pneumonia and influenza
* **Marcus Restrepo, MD, MSc, PhD** – American College of Chest Physicians (CHEST), Associate Program Director of Division of Pulmonary and Critical Care Medicine, UT San Antonio; Expertise: pneumonia diagnosis and care guidelines, pneumonia, infectious diseases
* **Gregory Rhunke, MD** – Associate Professor of Medicine; Expertise: hospital medicine, health outcomes research, pneumonia, diagnostic error
* **Lacee Sims** – patient experience (University of Utah)\*
* **Heath Young** – patient experience (University of Utah)\*

\*Note: both patient panelists also participated in three of four development-phase technical expert panels (TEP) for this measure, in addition to the final measure specification TEP

Study Team in Attendance:

* Barbara Jones
* Bahar Alimadadi
* Alec Chapman
* Tavis Huber
* Brittany Kent
* McKenna Nevers
* Lindsay Visnovsky
* Jian Ying

**Introduction**

The first technical expert panel (TEP) discussion for the final measure specification phase of measure development was held by video conference on Microsoft Teams on January 24, 2024. The TEP consisted of the seven panelists outlined above (4 clinicians, 1 PharmD researcher and health agency representative, 2 patients), in addition to the measure development team, led by Dr. Barbara Jones. Both patient panelists had previously participated in three of the development phase TEPs and the clinician panelists were all new participants to the TEP. The TEP agenda is included at the beginning of this report. Briefly, panelists and measure development team members introduced themselves and this was followed by a forty-minute presentation outlining the origin of the measure being developed and some of the measure’s performance characteristics, followed by an hour of panelist discussion. Prior to the formal beginning of panel discussion, TEP participants also shared real-time feedback on results as they were presented. At the beginning of the presentation, Dr. Jones oriented panelists to key feedback needs, namely: “a really strong vetting of the usability and actionability of these measures. So that means: is this [measure] something that, if implemented, if this measure were kind of rolled out [in current specification]… would this lead to an improvement in diagnosis and quality in pneumonia care? And how would that work? …would that be something that people can use to [improve]?” The meeting agenda sent prior to the TEP also focused on usability/actionability, impact, and the intersection of the current measure specification with anticipated changes in pneumonia patient diagnosis and care.