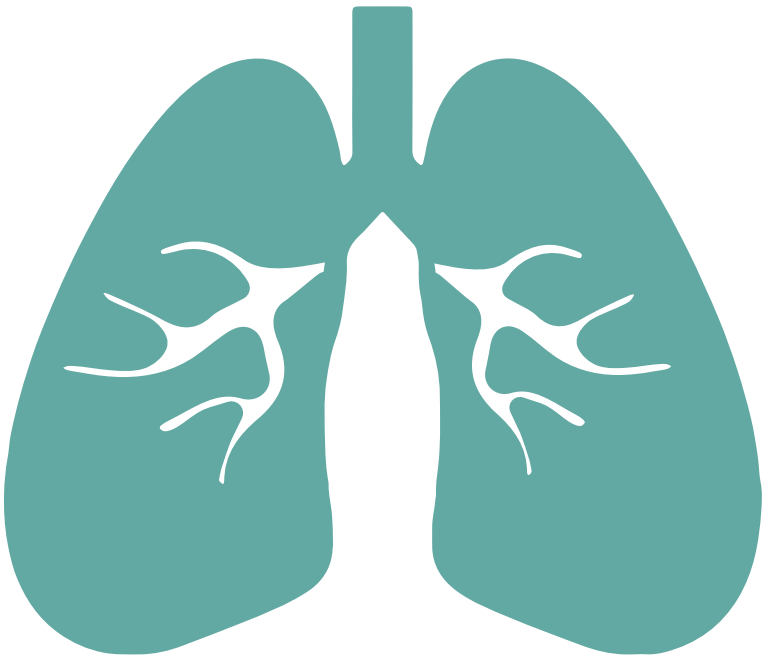
NQF Technical Expert Panel: Summary Report

Inappropriate Diagnosis of Community-acquired Pneumonia in Hospitalized Patients

December 2021



HMS
MICHIGAN HOSPITAL
MEDICINE SAFETY CONSORTIUM

TABLE OF CONTENTS

Introduction	3
Background	3
Technical Expert Panel	3
Discussion Summary	5
2.1 General Overview	5
2.2 Introduction of the CAP Measure	6
2.3 Validity of the CAP Measure	9
2.4 Reliability of the CAP Measure	11
2.5 Feasibility of the CAP Measure	13
2.6 Use of the CAP Measure	14
Responses to Issues Raised During the TEP	16

1 INTRODUCTION

Background

As part of preparation for submission of our measure on inappropriate diagnosis of community-acquired pneumonia (CAP), we convened a Technical Expert Panel (TEP) for input on measure specifications and naming. This report provides a summary of our TEP meetings, which were hosted from October 18 through October 28, 2021. Section 1 outlines the structure and composition of the group. Section 2 summarizes the presentation, member discussion and key findings. The discussion summaries presented are not meant to represent consensus, but rather to consolidate feedback. Finally, section 3 outlines the responses to issues raised during the TEP.

1.1 Technical Expert Panel

The TEP was comprised of 15 stakeholders with diverse perspectives and areas of expertise. The panel included experts from a variety of professional organizations, non-profit organizations, and governmental agencies (see table below).

Name, Credentials	Organizational Affiliation, City, State
David Newman-Toker, MD, PhD	Society to Improve Diagnosis in Medicine (SIDM), Baltimore, MD
Larissa May, MD, MSPH, MSHS	American College of Emergency Medicine (ACEP), Fair Oaks, CA
Teena Chopra, MD, MPH	Infectious Disease Society of America (IDSA), Detroit, MI
David Hyun, MD	Pew Research Center, Washington D.C.
Daniel Morgan, MD, MS	Society of Healthcare of America (SHEA), Baltimore, MD
Jason Pogue, PharmD, BCPS, BCIDP	Society of Infectious Diseases Pharmacists (SIDP), Plymouth, MI
David Baker, MD, MPH, FACP	The Joint Commission, Oakbrook Terrace, IL
Patty Gray, RN, CIC, FAPIC	Association for Professionals in Infection Control and Epidemiology (APIC), Scottsdale, AZ
Arjun Srinivasan, MD (CAPT, UHPHS)	Centers for Disease Control and Prevention (CDC), Atlanta, GA
Michael Liss, MD, MAS, FACS	American Urological Association (AUA), San Antonio, TX
Michael Pulia, MD, MS	Emergency Medicine; University of Wisconsin Madison, Middleton, WI
Peter Lindenauer, MD, MSc, MHM, FACP	Society of Hospital Medicine (SHM), Northampton, MA
Ella Kazerooni, MD, MS	American College of Radiology (ACR), Northville, MI
Marcos Restrepo, MD, MSc, PhD, FCCP	American College of Chest Physicians (CHEST), San Antonio, TX
Mark Metersky, MD, FCCP, FACP	American Thoracic Society (ATS), Avon, CT

Members of the TEP were invited to participate in three Zoom meetings (introduction, detailed discussion, response to questions) hosted on October 18, 20, and 25, 2021. All TEP members attended at least one meeting. In preparation for the meeting, the Michigan Hospital Medicine Safety Consortium (HMS) provided TEP members with an agenda and a background document providing project context and data on measure validity, reliability, feasibility, use, and usability.

The first hour of the October 18, 2021, meeting provided a general overview of the pneumonia measure that will be submitted to NQF. After this introductory meeting three were three separate small group meetings to have a detailed discussion of the over-diagnosis of pneumonia measure. These meetings were held October 18, 20, and 25th. We offered this meeting at three different times to best accommodate attendee schedules and allow for smaller groups to facilitate discussion. The TEP ended with a wrap-up session on October 28, 2021, a two-hour meeting during which we reviewed answers to questions raised during the detailed discussions and elicited final feedback. Each of these sessions were moderated by a physician associated with HMS.

2 DISCUSSION SUMMARY

This section summarizes feedback shared by TEP members during meetings hosted on October 18, 20, and 25, 2021. Each section focuses on the main components of each meeting:

Section 2.1 General overview
Section 2.2 Introduction of the CAP measure
Section 2.3 Validity of the CAP measure
Section 2.4 Reliability of the CAP measure
Section 2.5 Feasibility of the CAP measure
Section 2.6 Use of the CAP measure

The hour-long introduction on October 18, 2021, was attended by nine individuals. The three-hour content portion on October 18, 2021was attended by five individuals. The three-hour content portion on October 20, 2021, meeting was attended by six individuals. The three-hour content portion on October 25, 2021, pneumonia only session was attended by four individuals. Our closeout review session on October 28, 2021, was attended by 11 individuals.

2.1 Overview of HMS and the Initiative

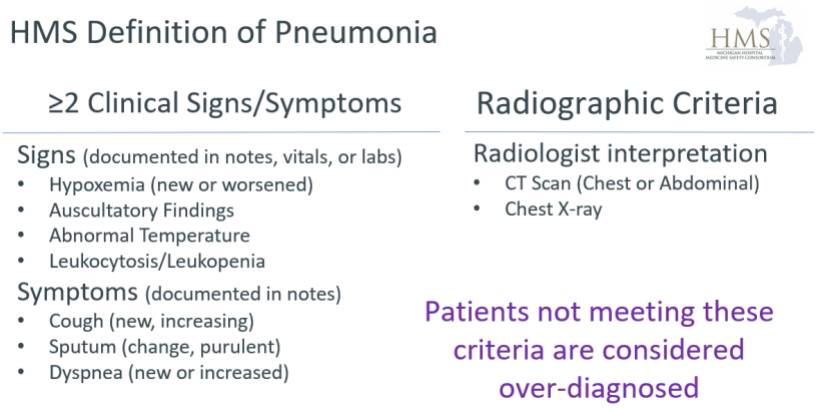
Before introducing and reviewing the specifics of the measures, the HMS team presented contextual background information. Content included the rationale for focusing on over-diagnosis, background on HMS, and how measures were created and validated.

First, the HMS moderator provided rationale for a focus on over-diagnosis. Several case examples were presented in which patients over-diagnosed with an infection they didn't have experienced negative outcomes because of their inappropriate diagnosis, including antibiotic-associated side effects and the implications of a missed or delayed true underlying diagnosis. A review of the literature pertaining to (1) diagnostic errors, (2) overuse of antibiotics, and (3) subsequent adverse events was also presented. Finally, existing measures and gaps in measurement were reviewed.

Second, the HMS moderator introduced HMS. HMS is a 62-hospital collaborative quality initiative (CQI) based in the state of Michigan and supported by the Blue Cross and Blue Shield of Michigan. The process for data collection was then outlined. HMS began collecting data on patients hospitalized for diagnosis of pneumonia in 2017; data were abstracted from medical records in a structured format by trained abstractors, who pseudo-randomly selected up to 8 cases for each measure every two weeks. The selection protocol was also outlined. The HMS dataset currently includes >35,000 pneumonia patients. Inclusion and exclusion criteria were reviewed, as was a summary of abstracted data (demographic data, antibiotic information, radiographic findings, microbiologic information, co-morbidities, vitals/oxygen/laboratory information, and symptoms).

Third, the moderator discussed the creation and use of measures of over-diagnosis of CAP and presented conceptual and operationalized definitions. For example, a conceptual definition for over-diagnosis of CAP was devised to identify patients treated for CAP who may

not have CAP. To operationalize the definition, the HMS team relied on national guidelines developed by professional societies. The operationalized definition of over-diagnosis of CAP was patients with ≥2 signs/symptoms of pneumonia on day 1/day 2 of hospitalization (or two days prior to hospitalization) AND having imaging findings consistent with pneumonia on days +/-3 of the hospitalization who were treated for CAP (and without a concomitant infection). Definitions of signs, symptoms, and radiographic findings were provided.



Data on pneumonia has been reported to HMS hospitals as quality measures since 2017. While hospitals are asked to review cases of over-diagnosis at their institutions to drive quality improvement, types of improvement initiatives implemented are at the discretion of the hospital.

This section wrapped up with HMS presenting the framework within which TEP members were asked to think about the rest of the information that would be presented over the following sessions.

- *Do you agree with the way that the measure was operationalized? Does it have "face validity"? What additional data/changes would you like to see?*
- *Do you believe lower measure scores represent better diagnosis? Better quality of care?*
- *Compared to other chart review measures, do you believe it is feasible for your hospital to collect these data?*
- *Could you envision this measure being used by / useful to your organization (e.g., guiding care improvement)?*
- *What perceived barriers do you see/foresee to using the over-diagnosis measure to guide care improvement? What unintended consequences do you see/foresee in using this measure?*

2.2 Introduction of the CAP Measure

HMS and the TEP members discussed how the pneumonia measure is defined and operationalized in regard to the inclusion and exclusion criteria used. Section 2.2.1 summarizes HMS' presentation, section 2.2.2 provides a summary of the TEP members' discussion, and section 2.2.3 summarizes the key takeaways.

Because we shared this data in conjunction with a second measure on urinary tract infections (UTI), we note that there is a high degree of overlap in a number of concepts for the UTI and pneumonia measures. In the 10/18/2021 and 10/20/2021 sessions, participants received an overview of both the UTI and pneumonia measures, with the UTI measure presented first. To utilize the TEP members' time most effectively, duplicative concepts were not repeated in the pneumonia session (e.g., data smoothing techniques). However, TEP members participating in the 10/25/2021 session only received pneumonia content, so an in-depth overview was presented to all participants. Therefore, this summary is based on the 10/25/2021 session. The summary of discussion represents all three sessions. Feedback specific to the UTI session is also available in that submission document.

2.2.1 Summary of Presentation

The HMS moderator laid the groundwork for the conversation by identifying the volume of patients currently treated for pneumonia; approximately 1 million adults in the United States. Approximately 50,000 individuals die from pneumonia each year, and among the elderly, hospitalization for pneumonia represents a greater risk of death compared to any of the other top 10 reasons for hospitalization. Finally, the HMS moderator noted that at least 1/8 of patients are over-diagnosed with pneumonia. These patients not only face harms associated from delayed diagnosis, such as longer lengths of stay, increased readmissions, and increased mortality, but are also treated unnecessarily with antibiotics that could result in harm. Indeed, 20% of hospitalized patients treated with unnecessary antibiotics develop an adverse drug event, most commonly gastrointestinal, renal, and hematologic. Patients receiving unnecessary antibiotics are also at risk of being treated for *C. difficile* infections.

The development of our definitions from a conceptual definition to an operationalized definition was reviewed in more depth. Time was spent reviewing the operationalized definition of community-acquired pneumonia. Patients must have a discharge diagnosis of pneumonia (ICD-10 code) and have received an antibiotic on day one or day two of the hospitalization, to exclude cases of hospital-acquired pneumonia. Next, because HMS wanted to ensure that antibiotic treatment could be attributed to antibiotic therapy and the clinical decision-making process, patients were not eligible for inclusion in the pneumonia cohort if they had a concomitant infection or had factors outside of clinician control – such as those leaving against medical advice. Those receiving brief empiric therapy (treatment <3 days) were also excluded. Finally, the HMS moderator outlined additional exclusion criteria (e.g., pregnancy/breastfeeding, immune compromised, and discharged with more than 14 days of antibiotics).

Over-diagnosis of CAP was defined using signs and symptoms and radiographic findings. To be classified as pneumonia, patients must have at least two signs or symptoms *and* radiographic findings from either a chest CT (or abdominal CT if it showed the lower lobes of the lungs) or a chest X-Ray. Signs included hypoxemia, auscultatory findings, abnormal temperature, leukocytosis/leukopenia. Symptoms included new or increasing cough, sputum (change/purulent); dyspnea (new increased). With regard to radiographic findings, abstractors were trained to look for terms including air space density/opacity/disease, consolidation, infiltrate; aspiration/aspiration pneumonia; abscess, cavitation, loculations; pneumonia, infection; ground glass, nodular airspace disease. The HMS team also indicated that the phrase "cannot rule

out pneumonia" is considered positive to give providers the benefit of the doubt. If a patient had a chest CT and a chest X-Ray within 24 hours of each other, the chest CT was prioritized. **Patients who did not meet these criteria were considered to be over-diagnosed with CAP.**

Finally, the HMS moderator reviewed the frequency with which these terms appeared in our dataset for both patients classified as pneumonia and over-diagnosis of pneumonia. The HMS moderator also noted that typically, over-diagnosis was due to symptoms in the absence of chest imaging findings. For example, only 23% (543/2,285) had positive X-rays (but insufficient symptoms).

HMS presented the following questions for discussion:

- *Do you agree with the way the measures' inclusion/exclusion criteria were operationalized? Does it have "face validity"? What additional data or changes would you like to see?*
- *Do you agree with the way the definition of over-diagnosis of pneumonia (PNA) was operationalized? Does it have "face validity"? What additional data or changes would you like to see?*

2.2.2 TEP Member Discussion

Concerns were raised about the measure name, specifically about the appropriateness of the term "over-diagnosis". One TEP member noted that misdiagnosis might be more appropriate than over-diagnosis. Another noted that false positive would be technically accurate but may not be easily interpreted by some audiences.

The conversation started with a discussion about screening for appropriate patients. HMS has used a discharge diagnostic ICD-10 code for pneumonia as an initial screen on which to apply inclusion and exclusion criteria to identify patients appropriate for the cohort. The TEP felt that this strategy was appropriate. There was conversation about whether screening should be based only on the primary diagnosis code rather than all discharge diagnosis codes; however, after discussing challenges associated with use of only the primary discharge diagnosis code (e.g., the primary code is often up coded to acute hypoxic respiratory failure or sepsis), it was largely agreed that use of all discharge diagnostic codes for cohort screening was appropriate.

There was also agreement that patients must receive antibiotic treatment on day one or day two of hospitalization in order to better ensure they were being treated for CAP rather than a hospital-acquired infection. That said, there was additional conversation surrounding classification of over-diagnosis should the patient only receive one or two doses of antibiotics, particularly by an emergency department provider. Data was presented regarding the frequency of antibiotic continuation on day three for patients over-diagnosed with pneumonia by an emergency medicine clinician (a vast majority) and we had ongoing conversation about including only patients who received $\geq$3 days of antibiotics. Part of this conversation took place in the context of using radiographic findings and signs/symptoms to determine pneumonia. TEP members discussed whether or not it is wrong (clinically) to start to treat a patient for pneumonia – prescribe antibiotics – even if they do not meet those criteria. Several TEP members noted if a

patient came in with a fever and showing other signs of pneumonia, without clear radiographic findings, they would treat that patient with antibiotics. The moderators also noted that HMS has purposely taken a very broad definition of radiographic evidence of pneumonia, including such phrases as "cannot rule out pneumonia" or "airspace density".

Finally, there was a conversation about defining fever differently in older adult patients, specifically noting that there are different potential cutoffs for this population. There was also concern that this would be of impact in congregant housing situations, such as skilled nursing facilities. The HMS team noted that this was something that was going to be looked into further.

There was also a conversation about the inclusion of medical patients only, as we wanted to focus on the least complex patients, so we excluded patients who were in the ICU, or being mechanically ventilated. One of the HMS moderators clarified that what we are looking at is not CAP, but CAP patients who are admitted to a general care service.

2.2.3 Key Takeaways

- HMS to review impact of using only primary ICD-10 discharge diagnostic codes versus all ICD-10 discharge diagnostic codes in patient screening.
    - Response: Data on primary vs. all discharge diagnostic codes were provided to the TEP during our follow-up session. They agreed not to change measure specifications (see below for details).
- HMS to review data further, including definition of fever and an evaluation of patients with only one sign or symptom.
    - Response: Data on fever and those with only one sign or symptom were compiled and presented to the TEP during the final session. There was general agreement not to adjust measure specifications as a result of these data (see below for details.
- The TEP members agreed that the measure title should be tested in workshops and focus groups to lessen any confusion regarding the measure purpose.
    - Response: potential titles were workshopped and then emailed back to the panel; the eventual final name for the measure was decided to be "inappropriate diagnosis of CAP" (see below for details).
- In general, the TEP members agreed that the current criteria for over-diagnosis of pneumonia were acceptable.
    - Response: Face Validity Confirmed

2.3 Validity of the CAP Measure

HMS and the TEP members discussed the validity of the pneumonia measure. Section 2.3.1 provides a summary of HMS's presentation, section 2.3.2 summarizes the TEP members' discussion, and section 2.3.3 summarizes the key takeaways.

2.3.1 Summary of Presentation

Following the discussion of how the HMS team defined over-diagnosis of CAP, we presented data on validity. We noted that we based our definition of CAP – and subsequently over-diagnosis of CAP – on national guidelines. Our definition is based on the 2007 IDSA/ATS consensus guidelines on the management of community-acquired pneumonia in adults as the diagnosis of CAP is based on the presence of select clinical features (e.g., cough, fever, sputum production, and pleuritic chest pain) and is supported by imaging of the lung, usually by chest radiography.

We next discussed the way HMS assesses patient-level validity using feedback from participating hospitals – hospitals receive site-specific data reports quarterly and they are encouraged to discuss patients that are classified as "over-diagnosed" with their local team. Hospitals that disagree with case classification are invited to bring that case back to the larger team for further evaluation, with measure updates considered based on feedback. However, due to the long-standing nature of this initiative, no recent modifications to the over-diagnosis measure have been made.

The HMS moderator also outlined a separate case assessment of patient-level validity conducted using physician implicit review. In 2020, the team conducted formal case reviews with 2 to 3 physicians, who were asked to review cases sampled from clinical scenarios where there may be ambiguity in case classification. Based on the feedback received, edits were made and now included in the final measure (e.g., abdominal CTs were added as a tool for diagnosis abnormal lung findings and chest CTs were considered the gold standard if obtained within 24 hours of either a chest X-Ray or an abdominal CT). The kappa for reviewer agreement was 0.72, indicating substantial agreement.

Next, the team presented data on which patients were at highest risk of over-diagnosis. Top predictors of over-diagnosis included dementia, altered mental status, being Hispanic or Black, and having a higher Charlson Comorbidity Index. We also presented data showing that hospitals that over-diagnose patients with pneumonia are also likely to over-diagnose patients with UTI (regression $R=0.53$, $p<0.001$).

The moderator wrapped up this portion of the presentation by reviewing the literature on diagnostic error and pneumonia. First, we noted that misdiagnosis of infections is common, representing nearly 15% of high-severity diagnostic error cases (resulting in serious, permanent disability, or death). Second, we noted that over-diagnosis of pneumonia is associated with use of antibiotics that are unnecessary and harmful. Third, we noted that discrepancy between admitting, and discharge diagnosis associated with 22.5% longer length of stay (~0.8 days longer). In the case of CAP, length of stay increase was more dramatic (~1 day).

This section concluded with a presentation of data on the effect of each unnecessary day of antibiotic use for *C. diff* infection, physician-reported adverse events, and patient-reported adverse events. Briefly the unadjusted odds ratio for these items are 0.94 (95% CI 0.79, 1.14), 0.98 (0.91, 1.05) and 1.05 (1.01, 1.09), respectively. Adjusted odds ratios were also presented, and patient-reported adverse events were again higher in the over-diagnosed cohort.

HMS presented the following questions for discussion:

- *Do you believe that lower measure scores represent better diagnosis? Better quality of care?*

2.3.2 TEP Member Discussion

One point of discussion among TEP members related to the ability to diagnosis pneumonia in patients with dementia or altered mental status, as they cannot reliably provide symptom data. Some TEP members felt the current inclusion criteria were reasonable because there are additional objective indicators within the inclusion criteria (e.g., hypoxia, abnormal WBC count, abnormal temperature).

There was also some concern about initial diagnosis and appropriateness of antibiotics. For example, if the admission diagnosis is sepsis, antibiotics were felt to be appropriate, whereas for pneumonia, antibiotics may be inappropriate. Additional concerns were raised about billing and coding – based on the payment and other items. The HMS moderators noted it is very hard retrospectively to distinguish between CAP and other similar presentations, such as a COPD exacerbation. One TEP member noted that establishing a measure is really a process of balancing the risks and benefits, which might be able to be addressed with empirical data and outcomes data – such as 30-day mortality, length of stay, etc.

The TEP panel members also discussed accuracy of radiology reports, and the possibility that radiologists over-call things that could be pneumonia. It was noted that we would rather the measure be overly sensitive as not to penalize clinicians who are acting with the information available at the time.

2.3.3 Key Takeaways

- The magnitude of harm of over-use vs. under-use of antibiotics needs to be demonstrated.
    - Response: Additional data included in NQF submission of harm of over-diagnosis and unnecessary antibiotic use.

2.4 Reliability of the CAP Measure

HMS and the TEP members discussed the reliability of the pneumonia measure. Section 2.4.1 summarizes HMS's presentation, section 2.4.2 summarizes the TEP members' discussion, and section 2.4.3 summarizes the key takeaways.

2.4.1 Summary of Presentation

The HMS moderator introduced the reliability section by discussing how data are collected. All data used in this measure are collected by trained data abstractors, the majority of whom have a nursing background. Each abstractor is provided with an extensive data dictionary

outlining established data protocols and charting the data that should be abstracted. These protocols and data include days of hospitalization for which to collect data (i.e., symptoms on the two days prior to admission and first two days of the hospital encounter), guidance on data entry (range/absolute value), location of data within the medical record (i.e., radiographic findings), definitions for all data elements, and instructions on how to handle missing data.

HMS also has an extensive audit process. Data from each hospital is audited at least once per year. The expectation is that hospitals would attain a ≥95% rate of accuracy. Since the inception of the project, over 300 cases have been audited. The moderator presented data on an additional audit conducted by the HMS program manager. The program manager selected 50 cases chronologically; representing 33 different hospitals. The results of this audit suggest that 95% of clinical findings were answered correctly, 92.3% of chest x-ray data was answered correctly, and 94.5% of chest CT data was answered correctly. As such, we identified an overall rate of abstraction accuracy of 93.7%.

Based on this audit, none of the 50 cases would have changed classification, meaning patients classified as pneumonia (or over-diagnosis of pneumonia) continued to be classified as pneumonia (or over-diagnosis of pneumonia) There was an inter-rater reliability of 1.0, which indicates strong to "almost perfect" reliability.

We also reviewed how reliable hospital estimates are at distinguishing variation between hospitals. The HMS moderator presented data on the percent of patients over-diagnosed by hospital and calculated an intraclass correlation of approximately 5%. This allowed us to compute the number of cases that hospitals participating in a measure would need to collect on an annual basis to reach a reliability threshold of 0.8 based on the Spearman Brown Prophecy; n=76. Based on the data currently being collected at HMS, we noted that the majority of hospitals would be able to meet an appropriate minimum threshold.

The reliability portion of the session concluded with a discussion of how HMS addresses small N sizes/data smoothing. In summary, HMS uses data for four quarters to better reflect the work that a hospital does over a one-year period. We then apply a mixed effect logistic regression model with a random intercept and random slope over time at the hospital level. Each hospital's adjusted rate reflects both change in performance over time and overall performance relative to all the other hospital averages. This model allows us to obtain more stable estimates for hospitals with few patients.

We then opened the meeting up for general conversation.

2.4.2 TEP Member Discussion

Conversation was limited as there was significant overlap between reliability of UTI and pneumonia.

2.4.3 Key Takeaways

- (from UTI insights) The group believed it might be necessary to move the data on the audits from the reliability section to the validity section.

    - Response: Pertinent data were moved to the validity section of the NQF submission
- There was consensus that hospitals would be able to collect a minimum threshold of data needed for an acceptable level of reliability (0.8).
    - Response: Evidence of Feasibility

2.5 Feasibility of the CAP Measure

HMS and the TEP members discussed the feasibility of the pneumonia measure. Section 2.5.1 provides a summary of HMS' presentation, section 2.5.2 summarizes the TEP members' discussion, and section 2.5.3 identifies key takeaways.

2.5.1 Summary of Presentation

The HMS moderator began a review of the feasibility section by noting that the HMS team evaluated several areas where we could pare down the data collected by HMS to streamline the data collection process. Exclusion criteria and frequency were reviewed in order to determine which, if any, could be removed without impacting outcomes. Based on this analysis, three items were removed from data collection that were infrequently captured in the dataset. These were Bandemia (12/34,626 cases), tree in bud (5/34,626 cases) and PaO2 (1/34,626). Next, we tested pared down data collection forms at five hospitals in Utah, which determined that reviewing cases for eligibility took one to three minutes and abstracting data took 15 minutes for easy cases to 30 minutes for more difficult cases. We also compared this to the amount of time it would take to abstract information for other NHSN measures requiring case review. We determined our abstraction timeframe was comparable to other measures. Finally, based on the Spearman Brown Prophecy data we presented earlier, we determined a threshold of abstracting 76 cases per year (standard threshold of 0.8) was reasonable, as 95.5% of the hospitals participating HMS were able meet this threshold. If the threshold was reduced to 0.7, all HMS hospitals met the metric.

The team also briefly reviewed if it was feasible to collect data electronically instead of via manual abstraction. To facilitate this process, we assessed if it would be feasible to electronically determine over-diagnosis of PNA by radiographs alone. If excluding symptoms from the definition, the percent of patients over-diagnosed with pneumonia changed, on average, by 6% per hospital (median 3.9%). If hospitals were placed in performance quartiles, this new definition would change the quartile for 64% of hospitals. As such, we determined the data for this measure still needed to be collected through manual abstraction.

HMS presented the following question for discussion:

- *Compared to other chart review measures, do you believe it would be feasible for your hospital to collect these data?*

2.5.2 TEP Member Discussion

There was conversation about whether electronic data capture was possible for this measure. There was discussion around the role of natural language processing (NLP) and

whether it would be possible to accurately capture this data. The HMS moderators re-reviewed the information on analysis of radiographic findings vs. symptoms and determined that use of NLP is not currently feasible, in particular due to the necessity of symptom extraction. The moderators also noted that we did meet with EHR vendors, who had no collective interest in making symptoms discrete elements in the medical record. One TEP member noted that if a hospital would deploy this measure, they could create a template for pneumonia symptoms.

In a discussion about number of cases needed, we reiterated that hospitals would need to abstract 5 to 6 cases per month. This led to a conversation about what percentage would we need to identify quality improvement. The HMS moderator noted that HMS identifies the top 25% as a cut off and then determine if that makes clinical sense. Future adjustments are then made based on that year's performance.

2.5.3 Key Takeaways

- There was general consensus that the measure was similar in difficulty to other chart review-based measures.
    - Response: Evidence of feasibility

2.6 Use of the CAP Measure

HMS and the TEP members discussed the use of the pneumonia measure. Section 2.6.1 summarizes HMS' presentation, section 2.6.2 provides a summary of the TEP members' discussion, and section 2.6.3 provides a summary of the key takeaways.

2.6.1 Summary of Presentation

The final section presented to the TEP was on use of the measure. We noted that while the over-diagnosis of pneumonia has NOT been an HMS pay-for-performance metric, these data have been reported to HMS hospitals since 2018. The HMS moderator also noted that during this time, improvement has been moderate (14% to 10% of all patients treated for pneumonia). It was also noted that this is expected to become an HMS pay-for-performance measure in 2023.

HMS presented the following questions for discussion:

- *Could you envision this measure being used by or useful to your organization? E.g., guiding care improvement.*
- *What perceived barriers do you see/foresee to using the overdiagnosis measure to guide care improvement? What unintended consequences do you see/foresee to using the over-diagnosis measure to guide care improvement?*

2.6.2 TEP Member Discussion

Overall, there was concern that people were already doing so well on this measure and whether there is enough room for growth/improvement. The group also noted that if we wanted

to put this measure forward, we needed to be very specific about the numerator and denominator definitions.

2.6.3 Key Takeaways

- There was mixed response to the use and usability. Some believed their organizations would be able to use this and would find it useful. Otherwise wanted more details.
    - Response: Details were added on outcomes of measure, outcomes after improvement in measure, and on measure specifications.
- Refine numerator and denominator definitions.
    - Response: Details were added to numerator and denominator definitions in measure specifications.

3. RESPONSES TO ISSUES RAISED DURING THE TEP

3.1 Responses to Issues Raised During the TEP

- Test name of measure in workshops and/or focus groups
    - HMS reviewed the name of our measure at two forums. The first was a standing meeting of the HMS team, which includes hospitalist, pulmonary/critical care, and infectious diseases physicians from Michigan Medicine. We also reviewed the data at a 2nd standing meeting, which allowed the team to review the measure name with a different group of physicians, including hospitalists, infectious diseases, geriatric/palliative care, internal medicine, and pediatric physicians. Between the two groups, we received feedback from 10 physicians. Based on the feedback we received, the use of word over-diagnosis was changed to inappropriate diagnosis. This name was then provided via email to the members of the TEP who had concern over the name; they agreed with the new name with one member noting that no name will be perfect for everyone. **Our new measure name is now inappropriate diagnosis of community-acquired pneumonia in hospitalized patients.**

- Evaluate fever cut-offs for adults
    - In the standard data analysis, HMS defines a fever as ≥38.1°C. In order to explore a lower fever threshold, we were able to leverage data from inpatients who had a temperature of 37.9 or 38.0°C. When we evaluated the number of patients that would change from over-diagnosis of pneumonia to CAP if we included these patients in this additional category, we found that only 9 patients would be impacted. Thus, we determined that changing the definition of fever was not necessary.

- Review impact of using primary ICD-10 codes vs. all ICD-10 discharge codes
    - The HMS standard of defining PNA was to include all patients who have an ICD-10 discharge diagnosis of pneumonia. However, this diagnosis does not have to be their primary/principal discharge diagnosis. As the first step in evaluating whether or not a change is warranted, we sought to determine how many patients had a primary discharge diagnosis that was not related to one of the following categories – pneumonia, respiratory failure, septicemia, or other lower respiratory disease. We determined that 12% (N=275) of cases of patients inappropriately diagnosed with CAP had a primary diagnosis that fell outside of these categories. We reviewed 9 medical records of these cases and felt the cases still reflected overdiagnosis of CAP. TEP members agreed with not changing measure specifications.

- Evaluate definition relative to patients with on one sign or symptom (not two)
    - The standard HMS definition for PNA is two or more signs/symptoms/lab values plus radiographic findings. We evaluated the percent of individuals that change from over-diagnosis to pneumonia if we looked at only one sign/symptom/lab value plus radiographic findings. To evaluate this, we did a

case review of 5 medical records in which patients had only one sign/symptom and radiographic findings. These cases were determined to accurately represent over-diagnosis of pneumonia. TEP members agreed with not changing measure specifications.

- Evaluate magnitude of harm of over-use vs. under-use of antibiotics
    - This is not something that we are directly able to do with our data, so we had to turn to the literature. Based on a literature review, we attempted to quantify the number of patients hospitalized with disease per year, the opportunity for improvement, number of patients who were harmed by antibiotic overuse, the number of patients harmed by underdiagnosis and the potential costs over/under diagnosis. We were able to document that there are approximately 1,000,000 pneumonia hospitalizations per year. Based on the percent of patients that we know are over-diagnosed from our own work – 1/8 of patients, we can extrapolate that 125,000 patients would be over-diagnosed annually, with an estimated 931,000 estimated unnecessary antibiotic days. Harm associated with over-diagnosis includes antibiotic associated adverse-events, antibiotic resistance, harm related to delayed diagnosis of congestive heart failure, myocardial infarction (MI) and lung cancer. Estimates suggest that at this rate, over-diagnosed patients would have a combined 37,634 adverse drug events and 4,794 cases of *C. diff*. **These data were added to the NQF submission.**