NQF questions:

This information is submitted on the NQF Testing Attachment

1) **Select the level of validity testing that was conducted**

   □ Patient or Encounter-Level (data element validity must address ALL critical data elements)

   □ Accountable Entity Level (e.g. hospitals, clinicians)

   □ Empirical validity testing

   □ Systematic assessment of face validity of performance measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

2) **For each level of testing checked above, describe the method of validity testing and what it tests**

Evaluation Criteria:
2b. Validity 2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) validity must be demonstrated for the data element level as well as for the computed performance score. For composite performance measures, validity must be demonstrated for the computed performance score by the time of endorsement maintenance; if empirical testing of the computed performance score is not feasible at the time of initial endorsement, acceptable alternatives include systematic assessment of content or face validity of the composite performance measure or demonstration that each of the component measures meet NQF subcriteria for validity (via either empirical testing or face validity) .

We performed validity testing on multiple levels and at multiple stages of measure development. A summary of validity testing is provided in the subsequent table with details provided in the following sections.

Table 1. Summary of Validity Testing

Process	Description (stage of measure)	Results	Interpretation
During Measure Development	*	*	*
A. Face Validity- National Guidelines	Based on National Guidelines and literature review **(Early Measure)**	IDSA/ATS CAP Guidelines[1,2]	Initial basis for definitions
B. Face Validity- Expert Feedback	Data Design and Publications Committee and Michigan Hospital Medicine Safety (HMS) Consortium Hospital Experts **(Early Measure AND Current Measure as Specified)**	Refined inclusion/exclusion criteria and measure specifications to current form	Measure refinement to current measure specifications
During Early Years (2017-2019) of Measure Use	*	*	*

Process	Description (stage of measure)	Results	Interpretation
C. Encounter-level Validity: Inappropriate Diagnosis Case Reporting	All inappropriate diagnosis cases reported to participating hospitals **(Early Measure AND Current Measure as Specified)**	Minor adjustments based on feedback from real cases and end-users	Minor measure refinement
During Late Years (2020-2021), Specific Measure Testing	*	*	*
D. Encounter-level Validity: Assessment of Effect of Abstraction Errors	Senior project manager reviewed data elements from 50 cases (representing 33 hospitals) to assess effect of any discrepancies on encounter-level validity **(Current Measure as Specified)**	Overall abstraction accuracy was 93.7%. No changes in inappropriate diagnosis classification due to discrepancies noted in audit	Encounter-level validity is high. Data abstraction is typically accurate; what mistakes are made generally do not affect case classification.
E. Encounter-level Validity: Structured Implicit Case Review	17 cases reviewed by 2-4 physicians to confirm classification **(Late Measure, only minor updates to measure after this assessment)**	The κ for reviewer agreement was 0.72	Indicates substantial agreement
F. Face Validity: Feedback from HMS hospitals (N=40 hospitals)	“Approximately, what percentage of cases called [inappropriate diagnosis of community acquired pneumonia (CAP)] by HMS do you agree are [inappropriately diagnosed] (0-100%)?” **(Current Measure as Specified)**	Median: 90% IQR: 80%-95%	Most participating hospitals believed the measure was highly accurate
G. Face Validity: National Expert Panel Feedback (N=14 experts)	Individuals form 14 national organizations participated in 2 week online technical expert panel (TEP) which involved discussion of measure. **(Current Measure as Specified)**	Generally, TEP members agreed with face validity. Additional questions/data requests were answered, and responses included below. Survey Question: “The inappropriate diagnosis of CAP measure as specified can be used to distinguish between better and worse quality hospitals.” Likert (1=Strongly disagree, 5=Strongly agree) 7/12 respondents (58.3%) reported that they agreed with this statement; 4/12 (33%) were neutral)	Additional feedback to improve utility of measure were provided and incorporated into the measure.
H. Face Validity: Patient Panel Feedback (N=7 patients)	Online focus group including 7 patients who had been hospitalized and treated for an infection **(Current Measure as Specified)**	Patients were asked what [inappropriate] diagnosis of infections meant to them and whether the measure would be valuable. They innately understood inappropriate diagnosis and its consequences.	Patients felt the inappropriate diagnosis of CAP measure was valid and important
I. Empirical Validity: Evaluated association with other measures of diagnostic quality	Evaluated association at hospital level between CAP inappropriate diagnosis and inappropriate diagnosis of UTI. **(Current Measure as Specified)**	Hospitals with higher rates of inappropriate diagnosis of CAP also had higher rates of inappropriate diagnosis of UTI; R=0.53 (i.e., moderate positive correlation)	Hospitals performing better on this measure were also better at appropriately diagnosing UTI

Process	Description (stage of measure)	Results	Interpretation
J. Empirical Validity: Evaluated association of inappropriate diagnosis of CAP with outcomes	Characterized antibiotic use in patients inappropriately diagnosed with CAP and the association of antibiotic use with adverse events after hospital discharge **(Current Measure as Specified)**	Median (IQR) 7 (5-9) unnecessary antibiotic days Each day of unnecessary antibiotic use increases odds (aOR: 1.05 [1.01, 1.08]) for developing a patient-reported antibiotic-associated adverse event after discharge.	Inappropriate diagnosis of CAP associated with unnecessary antibiotic use and antibiotic-related harm

*Cells intentionally left empty

Alt-Text for table 1:

Table 1 presents validity testing results and interpretation performed at various stages of measure development. Details are described in the text sections following the table.

A. Face Validity-National Guidelines

The inappropriate diagnosis of CAP measure was based on national guidelines for pneumonia and with addition expert feedback and review.

The 2009 Infectious Diseases Society of America/American Thoracic Society Consensus Guidelines on the Management of Community-Acquired Pneumonia in Adults define pneumonia as the following: “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.”[1] This definition is consistent with our measure which defines inappropriate diagnosis as any patient treated for CAP that is lacking clinical or radiographic criteria. We also evaluated symptom criteria from the Society for Healthcare Epidemiology of America’s evaluation of the use of non-specific symptoms in elderly populations.[3]

[1] Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007;44 Suppl 2:S27-72. doi:10.1086/511159. PCMID: PMC7107997.

[2] Metlay JP, Waterer GW, Long AC, et al. Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med*. 2019;200(7):e45-e67. doi:10.1164/rccm.201908-1581ST. PCMID: PMC6812437.

[3] Rowe, T., Jump, R., Andersen, B., et al. (2020). Reliability of nonlocalizing signs and symptoms as indicators of the presence of infection in nursing-home residents. *Infection Control & Hospital Epidemiology*, 1-10. doi:10.1017/ice.2020.1282

B. Face Validity-Expert Feedback

Throughout measure development, we obtained expert and stakeholder input via the following mechanisms:

1) Input from the Data, Design, and Publications (DDP) Committee of the Michigan Hospital Medicine Safety Consortium (HMS) early in measure development
2) Feedback from Experts in Quality, Antibiotic Stewardship, Diagnosis and Patient care from HMS hospitals

The **Data, Design, and Publications (DDP) Workgroup** was an ongoing meeting of champions and experts from HMS hospitals that met to address key issues related to measure methodology, including weighing the pros and cons of measure specifications, modeling, and use (e.g., defining the measure cohort and outcome) to ensure the measure was meaningful, useful, and well-designed. The group met approximately every 2 months during measure development and provided a forum for focused expert review and discussion of technical issues. They also provided final approval of the current submitted measure as specified.

List of DDP Workgroup Members:

- Suhasini Gudipati, MD Ascension Michigan St. Mary's Hospital
- Tina Percha, RN, MSN Beaumont Health
- Rajiv John, MD Beaumont Health
- Lama Hsaiky, PharmD Beaumont Health
- Priscila Bercea, MPH Beaumont Health Dearborn
- Scott Kaatz, DO Henry Ford Health System
- Allison Weinmann, MD Henry Ford Health System
- Emily Nerreter, MBA Henry Ford Health System
- Danielle Osterholzer, MD Hurley Medical Center
- Lisa Dumkow PharmD Mercy Health St. Mary's
- Anurag Malani, MD St. Joseph Mercy Ann Arbor Hospital
- Lakshmi Swaminathan, MD St. Joseph Mercy Ann Arbor Hospital
- Muhammad Nabeel, MD Sparrow Hospital
- Andrea White, PhD University of Utah Health
- Valerie Vaughn, MD, MSc University of Utah Health
- Vineet Chopra, MD, MSc University of Colorado Anschutz Medical Campus

Throughout measure development, we also provided opportunities from experts across the HMS collaborative to provide feedback. This included frontline clinicians, antibiotic stewards, quality improvement experts, c-suite members, and experts in quality measurement.

C. Assessment of Encounter-Level Validity: Inappropriate diagnosis Case Reporting

Once initial measure specifications had been agreed upon, we provided all inappropriate diagnosis cases to participating hospitals for review (N=2,301 cases of inappropriate diagnosis). Hospitals were encouraged to review these "fall-outs" with local experts in antibiotic stewardship, diagnosis, and quality as well as frontline clinicians to perform audit and feedback, identify trends, and assist with overall quality improvement. Occasionally, during this review the local team identified a potential issue with how the fall-out was determined based on the clinical scenario. In some instances, the case was reviewed, and we provided justification for considering the case inappropriately diagnosed. In other instances, modifications to the code and/or additional modifications to the data registry questions were required. Measure adjustments were more common during the initial launch of the measure (2017-2018). Since 2019, there have been no additional modifications to the measure based on this expert review. Since 2021, fall-out reporting has been based on the final submitted measure as currently specified.

D. Assessment of Encounter-Level Validity: Assessment of Effect of Abstraction Errors

To assess encounter-level data validity, the senior HMS project manager performed blind audits of 50 consecutive cases of patients with a diagnosis of CAP (appropriate or inappropriate). These cases included 33 hospitals. Cases were scored based correctness of data abstraction (1 point received if data element was answered correctly, 0 points if there was disagreement). The proportion of data elements abstracted correctly (based on the submitted measure as specified) were tabulated for clinical findings, chest x-ray findings, chest CT data, and overall abstraction accuracy. Correct data, as abstracted by the HMS project manager, were then reapplied to the measure definition to assess for changes in case classification.

E. Assessment of Encounter-Level Validity: Structured Implicit Case Review

In 2020, we conducted structured implicit review of cases of inappropriate diagnosis of CAP by 2-3 physicians to confirm accurate case categorization. Cases were randomly selected from “gray areas” that had been brought up during the initial measure development (e.g., patients with atelectasis as the only finding on chest imaging). During the review process, physician case reviewers had access to copies of medical record information such as diagnostic testing/results, emergency department note, history and physical note, progress notes, vital signs, and documented signs and symptoms. Reviewers were asked to independently assess whether they agreed with the classification of inappropriate diagnosis of CAP and whether they would empirically initiate antibiotics. If there was disagreement in classification, a discussion would commence that included ways to improve the measure to account for any errors in classification. We calculated the inter-rater agreement (prior to discussion) using **κ.** The comments generated through discussion were used as part of the feedback mechanism to improve the measure to the final specifications submitted here (edits in response to this feedback were minor, see details below).

F. Face Validity: Feedback from HMS hospitals (N=38 hospitals)

In October 2021 (after measure specifications had been finalized), we systematically assessed the perceived validity of the inappropriate diagnosis of CAP measure by soliciting feedback from all HMS hospitals. Via online survey, we asked all hospitals to answer the following question: “Approximately, what percentage of cases called [inappropriate diagnosis of CAP] by HMS do you agree are [inappropriately diagnosed] (0-100%)?”

G. Face Validity: National Expert Panel Feedback (N=14 experts)

Throughout measure development, we obtained expert and stakeholder input. In October 2021, we obtained formal expert feedback on the near final measure specifications by holding a two-week national technical expert panel (TEP) where societies and organizations who would potentially be impacted by the measure were asked to send a representative to provide feedback.

In alignment with the CMS Measures Management System guidance on TEP,[4] we convened a TEP to provide input and feedback from a group of recognized experts in relevant fields. To convene the TEP, we reached out to organizations whose members could potentially be impacted by the measure and asked them to nominate individuals for participation. We selected individuals to represent a range of perspectives, including Infectious Diseases physicians, pharmacists, pulmonologists, radiologists, hospitalists, emergency medicine physicians, regulatory agencies, as well as individuals with experience in quality improvement, performance measurement, diagnostic error, antibiotic stewardship, and health care quality. We held two weeks of structured TEP zoom calls consisting of a presentation of key issues, our proposed approach, and relevant data, followed by open discussion among TEP members. We solicited additional input and comments from the TEP via survey after the meeting. A summary of the TEP can be found in the **Appendix**.

Table 2. List of TEP Panelists and their Organizations

Organization/Institution	TEP Member
American College of Emergency Medicine (ACEP)	Larissa May
Centers for Disease Control and Prevention (CDC)	Arjun Srinivasan
Infectious Disease Society of America (IDSA)	Teena Chopra
Pew Research Center	David Hyun
Society for Healthcare Epidemiology of America (SHEA)	Dan Morgan
Society to Improve Diagnosis in Medicine (SIDM)	David Newman-Toker
Association for Professionals in Infection Control and Epidemiology (APIC)	Patty Gray
Society of Infectious Diseases Pharmacists (SIDP)	Jason Pogue
The Joint Commission	David Baker

Organization/Institution	TEP Member
Emergency Medicine Physician, University of Wisconsin	Michael Pulia
Society of Hospital Medicine (SHM)	Peter Lindenauer
American College of Radiology (ACR)	Ella Kazerooni
American College of Chest Physicians (CHEST)	Marcus Restrepo
American Thoracic Society (ATS)	Mark Metersky

Alt-text for Table 2: The fourteen TEP panelists and their organizations are listed.

Following the zoom expert panel, all participants completed an online survey that included questions related to validity, reliability, usability, etc. Related to measure validity, we asked TEP members:

a) How much do you agree/disagree with the following statement?
"The inappropriate diagnosis of CAP measure as specified can be used to distinguish between better and worse quality hospitals." 1=Strongly disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly agree.

b) Are there any key data elements you believe are missed or not accurately captured in the inappropriate diagnosis of CAP measure?

H. Face Validity: Patient Panel Feedback (N=7 patients)

To understand patient perspectives on the inappropriate diagnosis of CAP measure, we solicited patient feedback through a Patient Engagement Panel. This focus group was conducted on December 1, 2021 by the Community Collaboration and Engagement Team (CCET) which is part of the University of Utah Center for Clinical & Translational Science (CCTS). During this focus group, 7 patients and/or the caregivers of patients who had been hospitalized with infections were selected to provide feedback. Topics discussed included: how patients were diagnosed, what treatment they received, their understanding of risks and benefits with antibiotics, their perceptions about their illness and recovery, and how information about how hospitals diagnose and treat infections may inform their medical decisions. The discussion was guided by a Focus Group Discussion Guide (see Engagement Session Report for questions).

I. Empirical Validity: Evaluated association with other measures of diagnostic quality

To assess empirical validity for the inappropriate diagnosis of CAP measure, we identified and assessed the measure's correlation with other measures that target similar domains of quality for similar populations. The goal was to identify if better performance on this measure was related to better performance on other relevant structural or outcome measures. After literature review and consultations with measure experts in the field, there were very few measures identified that assess the same domains of quality.

To better understand whether inappropriate diagnosis is linked across conditions—and thus may reflect the general quality of diagnosis at a hospital—we assessed the association of inappropriate diagnosis of CAP with inappropriate diagnosis of UTI at the hospital level.

J. Empirical Validity: Evaluated association of inappropriate diagnosis of CAP with outcomes

We also assessed the association of inappropriate diagnosis with antibiotic-associated adverse events. First, we characterized antibiotic use in patients inappropriately diagnosed with CAP using descriptive statistics. Because duration was skewed, we report median (IQR/inter-quartile range) duration of antibiotic therapy.

[4] "CMS MMS Blueprint Supplemental Material: Technical Expert Panels." September 2021. https://www.cms.gov/files/document/blueprint-technical-expert-panels.pdf

Next, we evaluated the association of each day of unnecessary antibiotic therapy with patient outcomes at 30-days. Specifically, we were interested in the effect of each day of unnecessary antibiotic use on patient-reported antibiotic-associated adverse events (obtained through 30-day phone calls). We used generalized estimating equation models adjusted for patient characteristics to assess patient outcomes associated with each day of unnecessary antibiotic use.

2b.03 NQF question: Provide the statistical results from validity testing.

D. Encounter-level Validity: Assessment of Effect of Abstraction Errors

In 2021, 50 cases were chronologically selected for detailed audit. Audit findings were as follows:

Table 1. Results of detailed audit for data accuracy

Audit Elements	Results
Clinical Findings	95.7% of data elements abstracted correctly
Chest X-ray data	92.3% of data elements abstracted correctly
Chest CT data	94.5% of data elements abstracted correctly
Overall abstraction accuracy	93.7% of data elements abstracted correctly

Alt-text for Table 1. Results of detailed audit of clinical findings, chest X-ray, and chest CT data. Data accuracy ranged from 92.7% to 95.7%.

When errors found through the data audit were corrected, there were no changes in case classification, as shown in Table 2.

Table 2. Classification of cases in which audited data elements disagreed (n=50)

Abstractor Classification (original)	Auditor Classification (updated)	Number (n=50)
Inappropriate Diagnosis of CAP	Inappropriate Diagnosis of CAP	6
CAP	CAP	44
Inappropriate Diagnosis of CAP	CAP	0
CAP	Inappropriate Diagnosis of CAP	0

Alt-text for Table 2. When errors found through the audit were corrected (n=50 instances), there were no changes in case classification.

E. Encounter-level Validity: Structured Implicit Case Review

In 2020, 17 cases of inappropriate diagnosis of CAP underwent structured implicit case review by 2-4 physicians. **In 94% of cases (16/17) there was 100% agreement by reviewers that the cases represented inappropriate diagnosis**. In the remaining 6% (1/17) 1/3 reviewers agreed it was an inappropriate diagnosis. **The κ for reviewer agreement (prior to reconciliation) was 0.72** indicating substantial agreement. Of note, our case review involved "gray areas" rather than a random selection of cases. Thus, our true κ may be even higher. As a result of this case review process, we made minor refinements to our measure specifications including how chest CTs were assessed (they were given precedence over chest X-rays) and started including abdominal CTs with lung findings in the assessment/classification process.

F. Face Validity: Feedback from HMS hospitals (N=39 hospitals)

We systematically assessed the perceived validity (after finalization of measure specifications) of the inappropriate diagnosis of CAP measure by soliciting feedback from all participating HMS hospitals (N=39 hospitals) via the following question: “Approximately, what percentage of cases called ?PNA by HMS do you agree are ?PNA (0-100%).” Nearly all hospitals (97.4%, 38/39) responded. Respondents were local leaders or quality champions for the measures.

Median: 90% Inter-quartile range: 80%-95%

G. Face Validity: National Expert Panel Feedback

Based on conversations held during our two-week online TEP, the 14 national experts who attended our TEP generally agreed with the face validity and operationalization of the measure. They believed that patients we identified as being inappropriately diagnosed were, in fact, inappropriately diagnosed. The main concern brought up by panelists was a desire for more information on a balancing measure (i.e., under-diagnosis or missed diagnosis of CAP) and patient harm. There were also some concerns about the use of the word “over-diagnosis” in the measure name. As a result, we strengthened our literature review on under-diagnosis/missed diagnosis and added data on antibiotic overuse and patient harm as a result of inappropriate diagnosis of CAP. We also changed the measure name to “inappropriate diagnosis.” There were no changes to measure specifications suggested by the TEP.

TEP Survey results:

Table 3. Distribution of TEP responses to ***Question #1***: “*The inappropriate diagnosis of CAP measure as specified can be used to distinguish between better and worse quality hospitals*.”

Rating	# of Responses (N=12)	Percent (%)	Cumulative Percent (%)
5 (Strongly agree)	0	0	0
4 (Agree)	7	58.3%	58.3%
3 (Neutral)	4	33.3%	91.7%
2 (Disagree)	0	0	91.7%
1 (Strongly disagree)	1	8.3%	100.0%

Alt-text for Table 3: The majority (91.6%) of experts on the TEP responded “Agree” or Neutral” (7/12 and 4/23, respectively). There was one response of “Strongly disagree”.

Question #2. “What additional data would you like to see captured related to the inappropriate diagnosis of CAP? (free text)” N=14 respondents (free text question)

Table 4. TEP responses to ***Question #2.*** “*What additional data would you like to see captured related to the inappropriate diagnosis of CAP? (free text)*” N=14 respondents (free text question)

# of Responses N=14	Response	Our Action/Response to Comment
57% (8/14)	None or N/A	None. Confirmed validity of measurement.

# of Responses N=14	Response	Our Action/Response to Comment
14% (2/14)	Duration of Treatment	Added data on duration of treatment for patients inappropriately diagnosed with CAP to measure submission. **Patients inappropriately diagnosed with CAP received a median (IQR) 7 (5-9) antibiotic days, all of which were unnecessary.**
7% (1/14)	Balancing Measure	Added additional resources on studies of underdiagnosis to measure submission (see Evidence section)
7% (1/14)	Trend in Outcomes of Denominator Over Time as Inappropriate Diagnosis Decreases	Added data on outcomes over time to measure submission (see **Table 5**, below)
7% (1/14)	How many patients over 80 years old have only 1 sign or symptom.	Added data on those over 80 (see **Table 6**, below)

Alt-text for Table 5. The majority (57%) of experts on the TEP indicated that no additional data were needed. Suggestions for additional data included: a) duration of antibiotic treatment (2 panelists), b) balancing measure (1 panelist), c) trend in outcomes of denominator over time as inappropriate diagnosis decreases (1 panelist), and how many patients over 80 years old have only 1 sign or symptom (1 panelist). We addressed each of these in our measure submission.

Table 5. Trend in adverse outcomes over time as inappropriate diagnoses of CAP decreased from 2017 to 2020

Outcome	2017 (N=6405)	2020 (N=4961)
30-day Composite Outcome[a]	26.9% (1723)	25.4% (1260)
Death	3.5% (221)	2.9% (145)
Adverse Antibiotic Event	4.8% (306)	3.0% (147)

[a] Includes readmission, ED visit, death, *C. difficile*, and physician or patient reported antibiotic-associated adverse events

Alt-text for Table 5. From 2017 to 2020, there were decreases in the proportion of adverse outcomes including a 30-day composite outcome (includes readmission, ED visit, death, C. difficile, and physician or patient reported antibiotic-associated adverse events), death, and adverse antibiotic events.

Table 6. Comparison of inappropriate diagnosis and proportion of patients with only one sign or symptom in patients <80 Years vs patients age 80 or older

Age	All Patients	Inappropriate Dx	1 Symptom Only
<80	13,633	11.8% (1607)	2.3% (311)
≥ 80	4,960	14.0% (694)	3.6% (177)

Alt-text for Table 6. Table 6 compares the proportion of inappropriately diagnosed patients <80 vs ≥80 years with only 1 sign or symptom. The proportion of inappropriate diagnosis of CAP and having 1 sign or symptom only was greater in the 80 or older group (14.0% vs 11.8% and 3.6% vs 2.3%, respectively for inappropriate diagnosis of CAP and having only 1 sign or symptom).

H. Face Validity: Patient Panel Feedback:

A summary of the findings from the Patient Engagement Panel can be found in the **Appendix.**

Generally, the patients who participated in our panel innately understood the meaning of over-diagnosis or inappropriate diagnosis:

> **"[over-diagnosis is] taking a somewhat minor issue and overemphasizing it and then maybe overtreating it"**
>
> **"I was over-diagnosed by the doctor that I went to… I originally went because I had [a cough]… they didn't do any tests; he thought it was pneumonia and never did a test for it; he gave me 3 antibiotics within a 4-week time and so I feel like that is a perfect case of over-diagnosis. [Doctor says] hey, you're sick, I don't want to do a test, so take this." [Note. This participant was later admitted to another hospital with C. diff]**

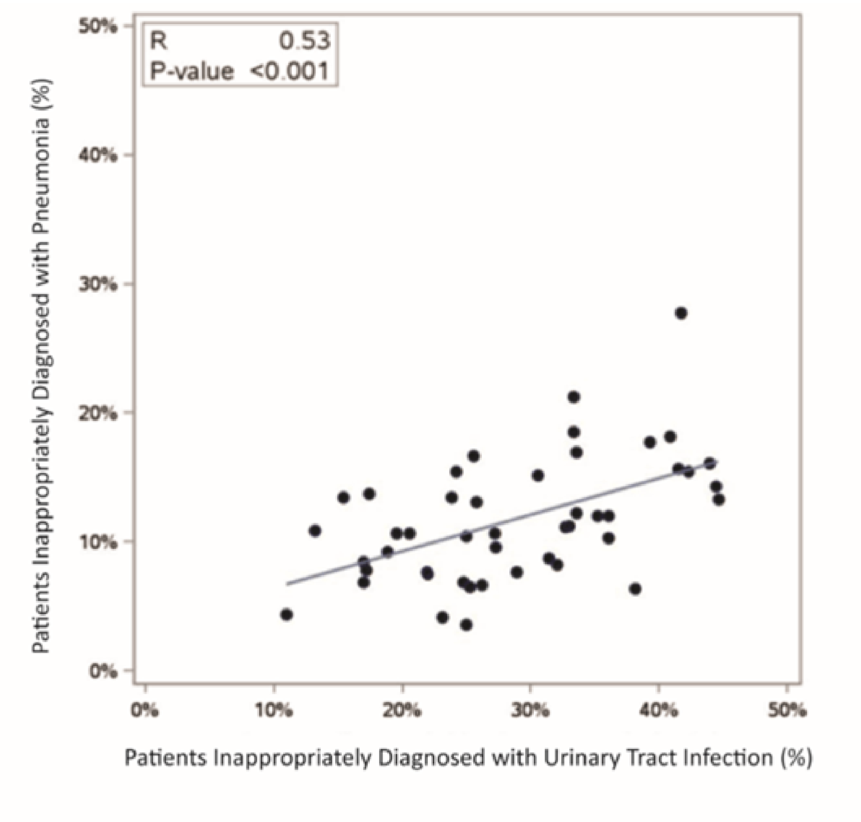
Patients also felt that measuring inappropriate diagnosis of infections was important and meaningful:

> **"That's [correct diagnosis] step 1… it takes me back to grad school…problem definition – you gotta make sure you're solving the right problem – that's the first step. If you don't, you're going to end up going down all these paths that are not going to lead you to the right answer."**
>
> **"If you were to have a measure of more correct diagnosis and incorrect diagnosis, and I would do it on the hospital scale, … I feel like if you were to get the correct diagnosis… I would automatically assume that you are getting the correct dose of medicine."**
>
> **"I would like it if they had a hospital rating… I think it would be beneficial, and I would really appreciate that. I feel that it would affect my decision of where I would go… it would definitely affect where I would guide my family or loved one to go."**
>
> **A participant has been looking for a care facility for his 98-year-old mother, utilizing U.S. News & Reports rankings. He said, "So yeah, I've been relying on that and I would definitely use something similar or look for something like that on the internet for a hospital."**

I. Empirical Validity: Association with Other Measures of Diagnostic Quality

To address whether inappropriate diagnosis of CAP was correlated with other domains of quality, we assessed whether inappropriate diagnosis of CAP (as currently specified) was related to the inappropriate diagnosis of UTI. This manuscript was published in *BMJ Quality & Safety*.[4] In it, we analyzed 10,398 patients treated for UTI and 14,085 patients treated for CAP from HMS hospitals between July 1, 2017 and March 31, 2020 and found that inappropriate diagnosis of CAP is moderately correlated with inappropriate diagnosis of UTI at the hospital level:

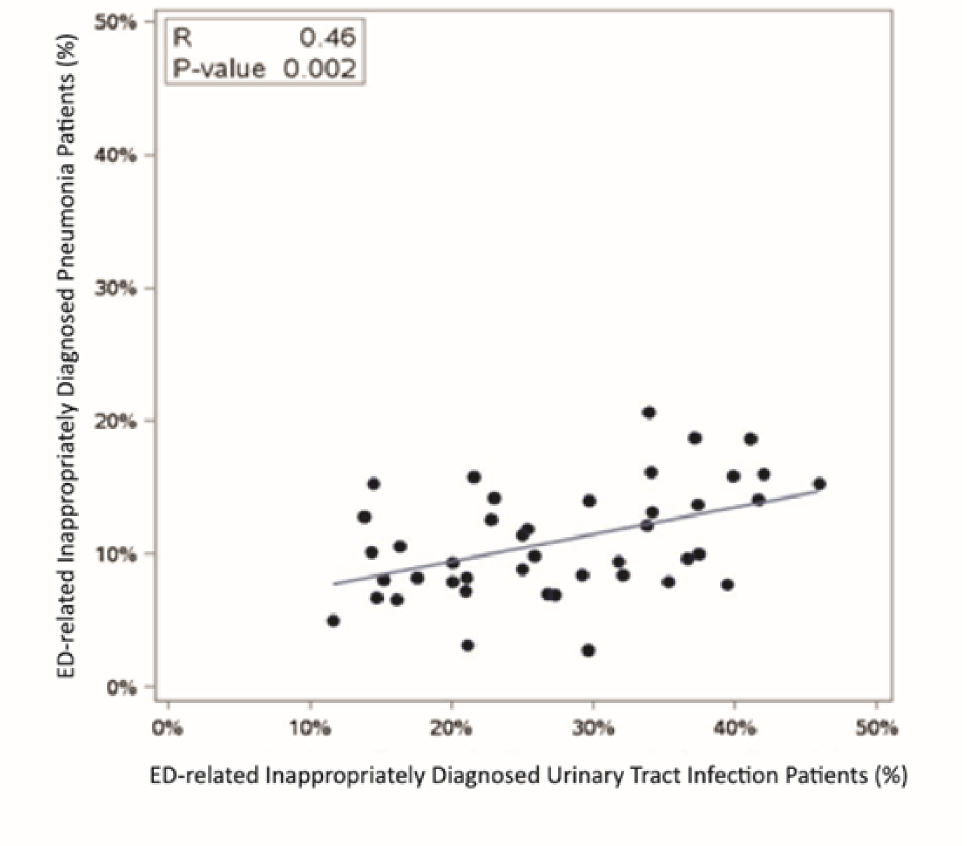
Figure 1. Relationship between inappropriate diagnosis of UTI and inappropriate diagnosis of CAP at the hospital level.



Alt-text for above figure: In a sample of 10,398 patients treated for UTI and 14,085 patients treated for CAP from HMS hospitals, the percent of patients with inappropriate diagnosis of UTI is moderately correlated with the percent of patients with inappropriate diagnosis of CAP at the hospital level (R=0.53; P<0.001).

These findings were also true for 2,049 patients initially inappropriately diagnosed in the Emergency Room.

Figure 2. Relationship between inappropriate diagnosis of UTI and inappropriate diagnosis of CAP in Emergency Rooms.



Alt-text for above figure: In a sample of 2,049 patients from 46 hospitals and diagnosed in the Emergency Room, the percent of patients with inappropriate diagnosis of UTI is moderately correlated with the percent of patients with inappropriate diagnosis of CAP at the hospital level (R=0.45; P<0.002).

[4] Gupta A, Petty L, Gandhi T, et al. Overdiagnosis of urinary tract infection linked to overdiagnosis of pneumonia: a multihospital cohort study. *BMJ Qual Saf*, 2022. doi:10.1136/bmjqs-2021-013565.

J. Empirical Validity: Association of Inappropriate diagnosis of CAP with Outcomes

There are three main harms associated with inappropriate diagnosis of CAP: delayed time to true diagnosis, antibiotic-associated adverse events, and antibiotic resistance. In our validation cohort of patients inappropriately diagnosed with CAP across HMS hospitals, patients inappropriately diagnosed with CAP received a median (IQR) 7 (5-9) antibiotic days, all of which were unnecessary. Those antibiotics were associated with harm such as antibiotic-associated adverse events, *C. difficile* infection, and antibiotic resistance without benefit (as they did not have bacterial infections). After adjustments, each additional day of antibiotic use in patients inappropriately diagnosed with CAP was associated with an increased odds ratio of 1.05 (1.01, 1.08) for developing a patient-reported antibiotic-associated adverse event.

Furthermore, as noted above in the response to TEP questions, we found that as inappropriate diagnosis of CAP (as currently specified) decreased over time, outcomes improved in HMS hospitals (**Table 5**, above).

2b.04 NQF Question: **Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)**

The validity of the inappropriate diagnosis of CAP measure is supported by three types of evidence: (1) strong face validity based on national guidelines and expert opinion and as gauged by feedback from Technical Expert Panel (TEP) members, patients, and end-users (hospitals, patients); (2) strong encounter level validity as demonstrated by implicit review, evaluation of data abstraction errors, and hospital encounter-level feedback; (3) external empiric comparisons with other quality measures; and (4) validity of the outcome.

Face validity

The validity of the measure is supported by strong face validity results, as measured by systematic feedback from the TEP. Perhaps even more important both patients and hospitals—the true end-users of the measure—found the measures to be valid. HMS hospitals who received measure scores found the measures to be highly valid, reporting they believed 90% of cases called inappropriate diagnosis of CAP were in fact inappropriately diagnosed.

Encounter-level Validity

Encounter-level validity is supported by substantial agreement between physician reviewers on case classification (κ=0.72) and by the long-standing general agreement by hospital experts with case classification during data feedback. Furthermore, in an assessment of the effect of abstraction errors on case classification, 93.7% of data elements were abstracted correctly and the minor discrepancies that existed resulted in no changes in case classification.

Empirical Validity Testing

The validity of the measure is further supported by the empiric validation results which demonstrate a correlation (in the expected strength and direction) between the inappropriate diagnosis of CAP measure and measures of inappropriate diagnosis of other infections, namely UTI. As expected, we found hospitals that performed worse on one measure also performed worse on the other. Thus, the inappropriate diagnosis of CAP measure may reflect the overall quality of diagnosis at a hospital.

Validity of the Outcome

The validity of the outcome is supported by the relationship between inappropriate diagnosis of CAP and and antibiotic-associated adverse events—including improvement in outcomes over time as measure performance improves.