NATIONAL QUALITY FORUM
Driving measurable health improvements together

# MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

**To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return**

## Brief Measure Information

**NQF #:** 3690
**Corresponding Measures:**
**Measure Title:** Inappropriate diagnosis of urinary tract infection (UTI) in hospitalized medical patients; Abbreviated form: Inappropriate diagnosis of UTI
**Measure Steward:** University of Michigan
**sp.02. Brief Description of Measure:** The inappropriate diagnosis of UTI in hospitalized medical patients (or "Inappropriate Diagnosis of UTI") measure is a process measure that evaluates the annual proportion of hospitalized adult medical patients treated for UTI who do not meet diagnostic criteria for UTI (thus are inappropriately diagnosed and overtreated).
**1b.01. Developer Rationale:** The goal of this measure is to improve the process for diagnosis and treatment of urinary tract infection (UTI). Literature has demonstrated that while UTI is one of the most common infectious etiologies for which patients are hospitalized, it is often inappropriately diagnosed, resulting in inappropriate antibiotic administration and delay in diagnosis of a true underlying condition. The implications of inappropriate antibiotics are well described and include risks of antibiotic-associated adverse events such as *Clostridioides difficile* infection, prolonged length of hospital stay, and antimicrobial resistance, all of which can increase patient morbidity and mortality. Missed or delayed diagnosis of a true underlying condition is equally troubling, as data suggest that diagnostic error results in the highest morbidity, mortality, and malpractice cost of any medical error. Through adoption of this measure, we anticipate a decrease in inappropriate diagnosis of UTI, a decrease in unnecessary antibiotic use, and improved patient outcomes.

---

**sp.12. Numerator Statement:** The measure quantifies adult, hospitalized medical patients inappropriately diagnosed with UTI. Here, inappropriate diagnosis is defined as patients treated with antibiotics for UTI who do not meet diagnostic criteria for UTI. Patients were considered inappropriately diagnosed if they received antibiotic therapy for a UTI but did not have at least one sign or symptom of a UTI.
**sp.14. Denominator Statement:** The denominator includes all adult, general care, immunocompetent, medical patients hospitalized and treated for UTI who do not have a concomitant infection.
**sp.16. Denominator Exclusions:**
**Exclusion Criteria:**

- Left against medical advice or refused medical care
- Admitted on hospice
- Pregnant or breastfeeding
- Spinal cord injury
- UTI-related complication (e.g., perinephric abscess)
    - Operationalized as >14 days of antibiotics at discharge

---

**Measure Type:** Process
**sp.28. Data Source:** Electronic Health Data
**sp.07. Level of Analysis:** Facility

## Preliminary Analysis: New Measure

### Criteria 1: Importance to Measure and Report

#### 1a. Evidence

**1a. Evidence.** The evidence requirements for a ***structure, process or intermediate outcome*** measure are that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

**The developer provides the following description for this measure:**

- This is a new process measure at the facility level that evaluates the annual proportion of hospitalized adult medical patients treated for urinary tract infection (UTI) who do not meet diagnostic criteria for UTI (thus are inappropriately diagnosed and overtreated).
- The developer provides a logic model that depicts the connection between patients inappropriately diagnosed with UTI and several negative health outcomes that can result, including a delayed or missed diagnosis of an unrelated underlying condition affecting the patient that might itself result in harm, as well as adverse events from administering the antimicrobial agents, and increasing antimicrobial resistance in the individual patient and in the patient's broader community.

**The developer provides the following evidence for this measure:**

- Systematic Review of the evidence specific to this measure? ☒ Yes ☐ No
- Quality, Quantity and Consistency of evidence provided? ☒ Yes ☐ No
- Evidence graded? ☒ Yes ☐ No

**Summary:**

- The developer highlighted recommendations from the May 2019 *Clinical Practice Guideline for the Management of Asymptomatic Bacteriuria: 2019 Update by the Infectious Diseases Society of America*, which issued a recommendation that across several varied categories of patients (healthy nonpregnant women, patients with diabetes, functionally impaired older persons or older persons in long-term care facilities, and patients with urethral catheters), none should be screened or treated for asymptomatic bacteriuria (ASB).
    - ASB is defined as "the presence of 1 or more species of bacteria growing in the urine at specified quantitative counts (>105 colony-forming units [CFU]/mL or >108 CFU/L), irrespective of the presence of pyuria, in the absence of signs or symptoms attributable to urinary tract infection (UTI)". The developer states the clinical guidelines match the submitted process measure in that a treatment for ASB would be consistent with a treatment for a patient that does not meet the diagnostic criteria for a UTI.
    - The recommendations were all rated as "strong", with evidence varying in quality from moderate to low based on the patient population.

- The recommendations each had at least 5 studies supporting them, with at least one randomized controlled trial (RCT) for all but one.
- The general recommendation to not treat or screen for ASB is because of high-quality evidence that antimicrobial treatments contribute to antimicrobial resistance
- The developer provided additional studies with evidence supporting the measure.
- One 2019, study found that inappropriate diagnosis of UTI is associated with a longer length of stay, of four days versus three days.
- A2017 study found evidence of higher readmission rates and mortality for patients inappropriate diagnosed with a UTI compared to patients not treated for ASB.
- The developer provided several studies supporting the harm associated with unnecessary antibiotic use, including a 2017 study that found that as many as 20 percent of patients receiving antibiotics experienced at least one antibiotic-associated adverse drug event

- The guideline that the developer cited did not provide an estimate of the benefit or consistency across the studies evaluated to develop the guideline, and the developer did not provide their own estimate of consistency across the collection of studies cited.

**Exception to evidence**

- N/A

***Questions for the Committee:***

- *Does the Committee agree there is sufficient evidence presented by the developer to link treatment for asymptomatic ASB (i.e. misdiagnosed UTI) to clinical outcomes, including the spread of drug-resistant bacterial infections, and adverse drug events?*

**Guidance from the Evidence Algorithm**

Measure does not assess a health outcome (Box 1) -> Evidence based on a systematic review (Box 3) -> Quality and Quantity provided, but not Consistency (Box 4) -> The evidence supports a strong recommendation (Box 6) -> Rate as MODERATE

**Preliminary rating for evidence:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

## 1b. Gap in Care/Opportunity for Improvement and Disparities

**1b. Performance Gap.** The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The measure was tested from 07/2017-03/31/2020 in 49 Michigan hospitals, finding 13,805 patients treated for a UTI, of whom 23.2% were inappropriately diagnosed.
- The developer reported hospital scores by performance decile:
  - In 2017, the median hospital in the best performing decile had 11.2% cases inappropriately diagnosed with a UTI. The worst performing decile had 53.8% of cases inappropriately diagnosed with a UTI.
  - In 2019, the median hospital in the best performing decile had 6.0% cases inappropriately diagnosed with a UTI. The worst performing decile had 31.7% of cases inappropriately diagnosed with a UTI.

**Disparities**

- In analyzing the demographics of the entire cohort, the developer found no differences in rates of inappropriate diagnosis by gender, race, or age, however, a significant difference was identified by payer.
    - Medicare patients were more likely to be inappropriately diagnosed than those with Medicaid or private insurance.

***Questions for the Committee:***

- *Is there a gap in care that warrants a national performance measure?*
- *Are there any concerns about the presence of disparities in this measure?*

**Preliminary rating for opportunity for improvement:** ☐ **High** ☒ **Moderate** ☐ **Low** **Insufficient**

## Committee Pre-evaluation Comments:

**1a. Evidence**

- Similar to the CAP measurement, the evidence shows adverse outcomes with treatment of asymptomatic bacteruria and there is no evidence presented on the validity of the these criteria in the diagnosis of UTI. In this case, the criteria are fairly sensitive and straight-forward and using the criteria to rule out UTI is probably appropriate. The concern would be in patients who are unable to provide symptoms.
- Agree with moderate rating, as evidence provided, while indirect, is compelling about over diagnosis of UTI and potential consequences
- process measure
- This is a process measure that evaluates the annual proportion of hospitalized adult medical patients treated for UTI who do not meet diagnostic criteria for pneumonia (thus are inappropriately diagnosed and treated). Using a logic model, the developer states there is a connection between inappropriately diagnosed UTI and negative patient outcome. A delayed or missed diagnosis of an unrelated underlying condition can cause patient harm, as well as adverse events from administering the antimicrobial agents, and increasing antimicrobial resistance for individual patients and within the communities. The developer provides a systematic review of relevant evidence, consistency of evidence, and grated evidence. The preliminary rating is moderate.
- Agree
- The evidence is strong against treating asymptomatic bacteriuria
- Where is the numerator statement? The denominator statement is unclear to me.
- The developer provided a SR, evaluated the evidence in the literature related to inappropriate diagnosis and treatment of UTI.

**1b. Gap in Care/Opportunity for Improvement and Disparities**

- Variability in performance in Michigan hospitals was presented.
- Significant spread in the data, suggesting improvement opportunities. Disparities by age observed, indicating this may be a more difficult population to diagnose or reflect more ingrained biases.
- moderate evidence moderate gap
- The measure was tested from 07/2017-03/31/2020 in 49 Michigan hospitals, finding 13,805 patients treated for a UTI, of whom 23.2% were inappropriately diagnosed. This demonstrates an opportunity for improvement. In analyzing the demographics of the entire cohort, the developer found no

differences in rates of inappropriate diagnosis by gender, race, or age; however, a significant difference was identified by payer, Medicare vs. Medicaid and private insurance.

- Medicare was more likely to be inaccurately diagnosed
- Significant performance gap shown.
- There is limited evidence that this measure will reduce inappropriate antibiotic prescribing.
- There is evidence of a performance gap and payer (Medicare) appears to play a role in disparate performance.

## Criteria 2: Scientific Acceptability of Measure Properties

**Complex measure evaluated by Scientific Methods Panel**? ☐ **Yes** ☒ **No**

**Evaluators:** Staff

### 2a. Reliability: Specifications and Testing

**2a1. Specifications** requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

**2a2. Reliability testing** demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

**Specifications:**

- Measure specifications are complex but clear and precise.

**Reliability Testing:**

- The data represents 13,805 hospitalized patients treated for UTI across 49 hospitals in the Michigan Hospital Medicine Safety Consortium (HMS) from July 1, 2017 – March 31, all of which were included in reliability and validity testing.
    - This dataset contained 82% academic hospitals, 82% metropolitan, 92% non-profit, and 69% hospitals with greater than 200 staffed beds.
- Reliability testing conducted at the Accountable Entity Level:
    - The developer performed a signal-to-noise analysis using a mixed-effect logistic model run as an empty model to calculate hospital variance (signal), within hospital variance (noise), and total variance, which were used to calculate the intraclass correlation coefficient (ICC).
        - Total Variance: 3.5151414
        - Hospital Variance (signal): 0.225271
        - Within Hospital Variance (noise): 3.28987
        - Based on signal-to-noise analysis, the developer calculated an ICC of 0.0641.
            - An ICC below 0.5 generally indicates poor reliability.
    - The ICC was used in the Spearman Brown formula to determine reliability for the entire hospital cohort using the median number of case abstracts and to determine the minimum case abstracts needed to achieve predetermined reliability thresholds (0.6, 0.7, 0.8, and 0.9).
        - After applying the median number of case extractions, the developed determined that reliability across the entire hospital cohort was 0.9.

        - Within a hospital cohort of 40 hospitals in 2019, 90 percent of hospitals were able to abstract the minimum 59 cases needed to achieve 0.8 reliability, and 95 percent of hospitals could abstract a minimum 35 cases to meet a 0.7 threshold.
- Reliability testing conducted at the Patient/Encounter Level:
    - Encounter-level validity was determined by assessment of effect of abstraction errors and structured implicit case reviews. The developer states that validity testing was conducted on all critical data elements, but since individual data element results were not provided in the submission and only the overall score was provided, NQF does not view this as sufficient to constitute complete patient/encounter level validity testing. It therefore is also insufficient for reliability testing at the patient/encounter level.

***Questions for the Committee regarding reliability:***

- *Do you have any concerns with the results of the signal-to-noise testing results that have been used to show reliability of the measure?*
- *Do you have any concerns that the lowest case abstraction quartile could not meet the lowest reliability threshold of 0.6?*
- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*

**Preliminary rating for reliability:** ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

**RATIONALE:**

Specifications are precise and unambiguous (Box 1) -> Reliability was conducted with the measure as specified (Box 2) -> Reliability testing conducted at the accountable entity level (Box 4) -> Signal-to-noise method used to determine reliability but unclear method used for calculating median number of case abstracts and ICC (Box 5) -> Unclear if empirical testing conducted on all critical data elements (Box 8) -> Rate as INSUFFICIENT

In signal-to-noise analysis, the internal variance is greater than the external variance and the intraclass correlation coefficient is well below 0.5, a range generally agreed to show poor reliability. It is not clear from the submission how applying the Spearman Brown prophecy formula leads to an overall reliability of 0.9. Additionally, only an overall score was provided for patient/encounter level reliability testing thus making it difficult to determine if all critical data elements were evaluated.

## 2b. Validity: Validity testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

**2b2. Validity testing** should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

**2b2-2b6. Potential threats to validity** should be assessed/addressed.

**Validity Testing**

- Validity testing conducted at the Patient/Encounter Level:
    - Encounter-level validity was determined by assessment of effect of abstraction errors and structured implicit case reviews. The developer states that validity testing was conducted on all critical data elements, but since individual data element results were not provided in the submission and only the overall score was provided, NQF does not view this as sufficient to constitute complete patient/encounter level testing.

- During the early years of measure use (2017-2019), the developer worked with participating hospitals to audit all cases of inappropriate diagnoses (N=3197).
- Using the current measure and data from 29 hospitals from 2020-2021, the senior HMS project manager audited 50 consecutive cases of patients diagnosed with UTI (appropriate or inappropriate).
    - The proportion of data elements that had been abstracted correctly was calculated. Inter-rater reliability was used to assess the different between the original case classification and the audit ("correct") case classification, with data errors identified. Overall data element abstraction accuracy was found to be 98.6%.
    - Only an overall score was provided for patient/encounter level validity testing thus making it difficult to determine if all critical data elements were evaluated
- Two to four physicians conducted a structured implicit case review of inappropriate diagnosis of UTI using 2020 data. Cases were randomly selected from among "gray areas" (i.e. less certain categorizations identified during initial measure development, such as patients with altered mental status). Reviewers independently assessed their level of agreement with the classification of inappropriate diagnoses of UTI. Minor updates were made to the current measure specifications following this review.

- Validity testing conducted at the Accountable Entity Level:
    - Face Validity
        - Face validity was assessed in 2021 using the current measure, receiving feedback from the 40 HMS hospitals used in the testing data, an 11-person technical expert panel (TEP), and patients.
            - Local leaders or quality champions from all 40 participating hospitals responded to the question, "Approximately, what percentage of cases called ASB by HMS do you agree are inappropriately diagnosed with ASB (0-100%)."
    - The median response was 90 percent, inter-quartile range was 80 – 97 percent.
            - The TEP was asked to respond to the following statement, "The inappropriate diagnosis of UTI measure as specified can be used to distinguish between better and worse quality hospitals," using a 5-point scale Likert scale.
            - Nine respondents agreed or strongly agreed. One respondent was neutral. One respondent strongly disagreed.
            - Face validity conducted with patients: The developer concluded from focus group discussions with patients and caregivers that this group understood the meaning of overdiagnosis and felt that measuring inappropriate diagnosis of UTI was both important and meaningful.
    - Empirical Validity of Measure Score
        - The developer tested whether better performance on this score was related to better performance on a measure of inappropriate diagnosis of community-acquired pneumonia (CAP). No suitable structural or outcome measures were found for this correlation so using a literature review and consulting with experts, the developer determined these two measure both represent signals of hospital quality that affect patient outcomes.
            - 24,483 patients from 46 hospitals were analyzed for this analysis, using data from 7/1/17-3/31/20. The developer found that inappropriate diagnoses of UTI has a moderate correlation with inappropriate diagnosis of CAP at the hospital level ($R=0.53$, $P<0.001$).

        - The findings were similar for the 2,049 patients initially inappropriately diagnosed within the emergency room setting (R=0.46, p=0.002).
    - The developer also assessed the association of inappropriate diagnosis with antibiotic-associated adverse events.
        - The developer found that in 2,733 hospitalized patients inappropriately diagnosed with UTI were treated with a median seven days of antibiotic therapy (IQR: 4-9 days), all of which were unnecessary.
        - The median length of stay for patients inappropriately diagnosed with UTI was 4 days and for those not treated with antibiotics, 3 days. The adjusted odds ratio for those treated with antibiotics was 1.37 (95% Confidence Interval (C.I.) 1.28-1.47, p<0.001).
        - Other associations showed potentially increased odds for those who received antibiotics (e.g. postdischarge mortality, postdischarge readmission, and discharge to post-acute care facility) but these associations were not statistically significant.
    - The developer next compared outcomes in patients inappropriately diagnosed with UTI vs. those who had ASB but were not treated with antibiotics using logistic generalized estimating equation models, inverse probability of treatment weighted by baseline covariates identified as significant, and other factors potentially associated with the outcome. Outcomes assessed included 30-day mortality, 30-day hospital readmission, 30-day emergency department visit, discharge to post-acute settings, Clostridioides difficile (C-Diff) infection at 30 days, and duration of hospitalization after urine testing.

**Exclusions**

- The developer was able to estimate the percent of exclusions from the literature and directly from the HMS dataset. Total exclusions were estimated at 1.8%-2.47%.
- Exclusion criteria was not tested. The criteria were provided to participating hospitals and the TEP to ensure they appeared feasible and reasonable. Some TEP members suggested additional populations to include in the future but believed that starting with a less contentious group (i.e., hospitalized medical patients) would be a good start before moving into more difficult populations (e.g., nursing homes).

**Meaningful Differences**

- Hospitals in the top 10$^{th}$ percentile for performance have approximately 12 fewer patients inappropriately diagnosed with UTI per 100 patients treated for UTI than the median. Those in the lowest 90$^{th}$ percentile for performance have about 15 more patients inappropriately diagnosed per 100 patients treated for UTI than the median. Hospitals were also analyzed by quartiles and the developer found a statistically significant difference in performance between all adjacent quartiles.

**Missing Data**

- The developer states that the percentage of encounters with missing data (labeled “unknown” or “not available” in the abstraction tool) was 5.2 percent (714/13,805). The majority of missing demographic data was for the ethnicity category, missing 14.6 percent (2,019/13,805). The majority of missing data relevant to UTI was for urinary catheter, missing 3.5 percent (484/13,805). The developer states that most missing data did not exist in the medical record and was not due to errors in abstraction.

***Questions for the Committee regarding validity:***

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*

**Preliminary rating for validity:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

## Committee Pre-evaluation Comments:

**2a. Reliability**

- 2a1. Reliability-Specifications
  - I am not sure the hospital level variance vs within hospital variation is probably not the most relevant measure of reliability. It is unclear to me how the calculated within hospital variability.
  - Seems very unclear, and the data suggest some very low levels of reliability. This is often challenging to discern in acute care settings
  - insufficient evidence
  - No concerns.
  - showed poor reliability
  - I agree with staff that reliability and validity testing is not sufficient.
  - Yes. The data are quite limited in this factor.
  - Same concerns as the prior measure on CAP
- 2a2. Reliability – Testing
  - Reliability of reporting and documenting symptoms is going to vary and may not be reliable across different hospitals or different service lines, or different patient populations
  - Yes
  - insufficient evidence
  - Preliminary rating is insufficient.
  - yes
  - see above comment
  - As above.
  - Same concerns as the prior measure on CAP

**2b. Validity**

- The clinical validity in ruling out UTI is reasonable. The measurement validity would be challenging in patients who are unable to provide symptoms. Those should be excluded. Also anyone without a positive urine culture is excluded. Patients may be treated without having a urine culture sent so this metric will undercount inappropriate treatment
- Agree with moderate rating, face validity is sound and important to patients. Appears related to hosptial quality as measures correlated with CAP misdiagnosis rate.
- moderate evidence
- Both face validity and empirical validity were at the Accountable Entity Level. There appears to be no major concerns.
- no concerns

- seems reasonable
- No.
- no concerns

**2b2-2b6. Potential threats to validity**

- 2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)
    - Patients who are unable to describe symptoms should be excluded. Some stratified or risk-adjustment should be conducted using age and gender and chronic Foley use as a minimum
    - Exclusions appropriate and do not seem to be high %
    - ??
    - No risk adjustment.
    - not risk adjusted
    - N/A
    - No risk adjustment. Only a small fraction of patients was excluded.
    - There is no risk-adjustment
- 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)
    - Not treating patients without clinical symptoms of UTI would be a meaningful quality and clinical outcome
    - 3-5%, minor concern for missing data
    - no
    - Exclusion is estimated about 1.8-2.5%, which seems small. The number of missing data does not seem to be significant, about 5% overall. Preliminary rating for validity is moderate.
    - no concerns
    - I worry that the definition does not distinguish between baseline vs. new symptoms, many older adults may have urinary frequency or urgency at baseline, how will those be handled? Also the definition involves urine culture which usually takes 2-3 days to result, so if someone is treated with abx but later stopped upon negative culture result, how will that be counted?
    - Not covered.
    - Some data was missing from the EHR, not missing from abstraction

## Criterion 3. Feasibility

**3. Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The measure uses chart-abstracted data routinely collected during the normal process of patient care; no additional data are required. In the studied hospital cohort, the developer reported just 5.2% of encounters had missing data, with “little effect” on whether the case could be classified as an inappropriate UTI diagnosis.
- Some data elements needed to calculate the measure must be chart-abstracted, and the developer found the measure was not feasible to transition to an eCQM.
- The data elements that must be abstracted are the symptoms of UTI, which are generally documented in the medical record in free text, with locations that vary based on the medical record and site-specific implementation factors.

- The minimum cases that need to be abstracted in order to meet a reliability threshold of .7 is just 35 cases/year, which 95% of studied hospitals were able to meet. To meet a reliability threshold of .8, 59 cases must be sampled, which 90% of studied hospitals were able to do.
- The developers surveyed studied hospitals, only 22.5% of whom reported it was "difficult" or "very difficult" to collect the needed data.

***Questions for the Committee:***

- *Is the Committee confident that the experience of the hospital cohort studied by the developer is broadly representative of hospitals which may report this measure nationwide?*

**Preliminary rating for feasibility:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

## Committee Pre-evaluation Comments:

**3. Feasibility**

- This metric involves chart abstraction for 59 charts a year. Many elements could be easily obtained by chart review (e.g. fever), the other sx would be difficult to obtain
- Abstraction seems very labor intensive
- moderate - none
- Over 90% of surveyed hospitals were able to meet a reliability threshold of 0.7 or 0.8. And about 22.5% of hospitals reported having difficulty in collecting the needed data. Preliminary rating is moderate.
- UTI diagnosis fields are not specific to UTI, symptoms or diagnosis
- There is added burden of chart abstraction.
- Data collection was difficult in a large fraction of cases.
- It seems that most of the hospitals were able to do the data collection

## Criterion 4: Use and Usability

### 4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

**4a. Use** evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

**4a.1. Accountability and Transparency.** Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

**Current uses of the measure**

| | | | | | | |
|---|---|---|---|---|---|---|
| Publicly reported? | ☐ | Yes | ☒ | No | | |
| Current use in an accountability program? | ☒ | Yes | ☐ | No | ☐ | UNCLEAR |
| Planned use in an accountability program? | ☒ | Yes | ☐ | No | ☐ | NA |

**Accountability program details**

- The measure is currently used in an accountability program, as of January 1, 2018. Blue Cross Blue Shield of Michigan sponsors the Michigan Hospital Medicine Safety Consortium (HMS) and awards financial incentives based on performance on this measure. The program also benchmarks hospitals against their own prior performance as well as other hospitals in the program.
- Additional planned uses include public reporting and public/health disease surveillance.

**4a.2. Feedback on the measure by those being measured or others.** Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

**Feedback on the measure by those being measured or others**

- Hospitals participating in HMS were given data on their performance relative to other hospitals, with specific lists of each patient that was considered inappropriately diagnosed with a UTI so that hospitals could review those cases. Hospitals then provided case-specific feedback back to the developer.
- The developer also sought open-ended feedback on the measure from participating hospitals, as well as asking about specific barriers to using the measure. In addition, the developer conducted a patient engagement panel, and no concerns were raised about the measure from patients.
- The developer updated the measure to reduce the number of cases that needed to be abstracted to obtain a reliable measure result and reduced the number of data elements needed.

***Questions for the Committee**:*

- *Has the measure been sufficiently vetted in real-world settings by those being measured or others?*

**Preliminary rating for Use:** ☒ **Pass** ☐ **No Pass**

## 4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

**4b. Usability** evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

**4b.1 Improvement.** Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

**Improvement results**

- Since launching this measure in the HMS accountability program in 2017, the developer observed a 37% decrease ($p<0.001$) in the percentage of patients inappropriately diagnosed for UTI, when measured against results in the first quarter of 2020. This shows substantial improvement attributable to the accountability program.

**4b2. Benefits vs. harms.** Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

**Unexpected findings (positive or negative) during implementation**

- There were no unexpected findings identified by the developer.

**Potential harms**

- The developer reported that only 25% of hospitals in the cohort foresaw possible unintended consequences from the implementation of this measure, including possible delays in administration of antibiotics for patients who have a UTI and patients dissatisfied with not receiving antibiotics.

***Questions for the Committee**:*

- *Can the performance results be used to continue to improve performance in other accountability applications beyond those presented here?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

**Preliminary rating for Usability:** ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

## Committee Pre-evaluation Comments:

**4a. Use**

- only reported in Michigan currently
- Used for internal QI data provided back to hospitals by payers. Planned use for public reporting
- yes
- This measure is not currently public reported. But the measure is being used in accountability programs. Blue Cross Blue Shield of Michigan sponsors the Michigan Hospital Medicine Safety Consortium (HMS). It is also used as an external benchmark program to evaluate hospital performance.
- no concerns
- N/A
- This measure has seen limited use.
- Feedback was considered.

**4b. Usability**

- unintended consequences of delayed antibiotics and sepsis should be measured.
- I think concerns about harms are minimal given clear benefits.
- moderate
- Since launching this measure in the HMS accountability program in 2017, the developer observed a 37% decrease ($p<0.001$) in the percentage of patients inappropriately diagnosed for UTI, when measured against results in the first quarter of 2020. This shows substantial improvement attributable to the accountability program. Usability is rated as moderate.
- none
- no concerns
- Harm could come from delayed diagnosis, but this seems small compared to potential improvements in diagnosis.
- No concerns about usability

## Criterion 5: Related and Competing Measures

**Related measures**

- Two measures were identified as related:
  - NQF#0138: National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure
  - NQF#0684: Percent of Residents with a Urinary Tract Infection (Long Stay)

**Harmonization**

- N/A

# Committee Pre-evaluation Comments:

**5: Related and Competing Measures**

- n/a
- Seems different from CAUTI measure and long-stay UTI measure. Has distinct value.
- 2 measures related
- Two measures were identified as related: NQF#0138: National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure; NQF#0684: Percent of Residents with a Urinary Tract Infection (Long Stay).
- no
- Neither of the two measures listed appear to compete.
- no concerns

## Public and NQF Member Comments (Submitted as of June 10, 2022)

**Comments**

### Comment 1 by: Submitted by Valerie Vaughn, Michigan Hospital Medicine Safety Consortium *(#3690 measure developer)*

This public comment is to address concerns about reliability and validity testing at the critical data element level. We did not include data element validity testing in the original submission but rather reported encounter level validity. We also have data element validity available and include it here: SUMMARY: Critical data element validity testing was conducted by a senior project manager who reviewed all critical data elements from 50 abstracted cases (representing 33 hospitals). Overall, the percent agreement for abstractor and auditor for critical data elements for signs/symptoms of UTI ranged from 94% to 100%. This suggests that data element validity is high and adds to our already submitted information that encounter level validity is high. DETAILS: Critical data elements for clinical signs/symptoms of UTI were examined by the senior project manager in blind audits of 50 consecutive patients with a diagnosis of UTI (appropriate or inappropriate) from 33 hospitals. Data elements 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 cases in which there was agreement for each data element were tabulated for clinical signs/symptoms of UTI and overall abstraction accuracy. Audit findings were as follows: Signs/Symptoms of UTI: Percent agreement between abstractor and auditor for critical data elements: Urgency 100% Rigors 98% Frequency 96% Dysuria 94% Suprapubic Pain or Tenderness 96% Acute Hematuria 94% Costovertebral or Flank Pain Tenderness 100% Fever (>38°C) 98%

Altered Mental Status 96% Temperature >38.0 98% Temperature <36.0 98% Heart Rate >90 BPM 96% Respiratory Rate >20 br/min 98% White blood count >10K/µL 98% Hypotension (SBP < 90 mmHg) 96%

**Comment 2 by: Submitted by Valerie Vaughn, Michigan Hospital Medicine Safety Consortium (*#3690 measure developer)***

This public comment is to address concerns about reliability testing at the accountable entitle level. There are concerns that our ICC appears low (0.0641). We would like to clarity that the ICC of 0.0641 applies only if a single case were obtained from each hospital. This indicates that if each hospital performed 1 case abstraction, there would be high variability and poor reliability. However, we do not suggest each hospital only conduct 1 case abstraction. The Spearman Brown Prophecy provides an estimation of reliability after adjusting the number of measurements. When the median number of case counts for the entire cohort (N=133 median cases per hospital in measure development hospitals) is applied to the Spearman Brown formula, the overall reliability was 0.901 (well above the 0.5 threshold noted for "poor reliability"). The 0.901 was calculated as follows: Median case abstractions: 133 (IQR 92-154) Reliability or ICC for 133 cases (i.e., ICC/reliability for a typical HMS hospital): (133*0.0641)/(1+(133-1)*0.0641)=0.901 Through this same calculation, using the Spearman Brown Prophecy, we calculated the number of annual cases needed to achieve each reliability threshold: Reliability---Number of annual cases needed 0.6---22 0.7---35 0.8 (standard)---59 0.9---132 Thus, we attain reliability of 0.8 (standard reliability for a quality metric of this stakes) with 59 cases per hospital which is our suggested target number of cases for the measure.

## Scientific Acceptability Evaluation

### RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   **Submission document:** Items sp.01-sp.30

   ***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**
   - No concerns.

### RELIABILITY: TESTING

3. **Reliability testing level**

   ☒ **Accountable-Entity Level** ☒ **Patient/Encounter Level** ☐ **Neither**

4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**

   ☒ **Yes** ☐ **No**

5. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   **Submission document:** Question 2a.10

- Reliability testing was conducted using data from 49 hospitals in the Michigan Hospital Medicine Safety Consortium (HMS) from 07/01/2017 – 03/31/2020
- The developer stated they conducted reliability testing at both the patient/encounter level and accountable/entity level
- Accountable Entity Level Testing
    - The developer performed a signal-to-noise analysis, then calculated hospital variance (signal), within hospital variance (noise) and total variance, which were used to calculate the intraclass correlation coefficient (ICC).
    - The ICC was used in the Spearman Brown formula to determine reliability for the entire hospital cohort using the median number of case abstracts and to determine the minimum case abstracts needed to achieve predetermined reliability thresholds (0.6, 0.7, 0.8, and 0.9).
- The developer stated that all critical data elements were tested in validity testing but only provided an overall score. In NQF's assessment this does not show validity of all critical data elements and thus cannot be used to demonstrate reliability.

7. **Assess the results of reliability testing**

   **Submission document:** Question 2a.11

    - Accountable Entity Level Testing
        - Based on signal-to-noise analysis, the developer calculated an ICC of 0.0641. After applying the median number of case extractions, the developed determined that reliability across the entire hospital cohort was 0.9, which was considered to be strong and meeting the pre-determined threshold for reliability for high stakes measures.
        - Within a hospital cohort of 40 hospitals in 2019, 90 percent of hospitals were able to abstract the minimum 59 cases needed to achieve 0.8 reliability, and 95 percent of hospitals could abstract a minimum 35 cases to meet a 0.7 threshold

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? **NOTE:** If multiple methods used, at least one must be appropriate.

   **Submission document:** Question 2a.10-12

   ☒ **Yes**

   ☒ **No**

   ☐ **Not applicable**

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   **Submission document:** Question 2a.10-12

   ☐ **Yes**

   ☒ **No**

   ☐ **Not applicable** (patient/encounter level testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☐ **High** (NOTE: Can be HIGH only if accountable-entity level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity level testing has not been conducted)

☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

☒ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - Specifications are precise and unambiguous (Box 1) -> Reliability was conducted with the measure as specified (Box 2) -> Reliability testing conducted at the accountable entity level (Box 4) -> Signal-to-noise method used to determine reliability but unclear method used for calculating median number of case abstracts and ICC (Box 5) -> Unclear if empirical testing conducted on all critical data elements (Box 8) -> Rate as INSUFFICIENT
    - In signal-to-noise analysis, the internal variance is greater than the external variance and the intraclass correlation coefficient is well below 0.5, a range generally agreed to show poor reliability. It is not clear from the submission how applying the Spearman Brown prophecy formula leads to an overall reliability of 0.9. Additionally, only an overall score was provided for patient/encounter level reliability testing thus making it difficult to determine if all critical data elements were evaluated.

### VALIDITY: TESTING

12. **Validity testing level (check all that apply):**

    ☒ **Accountable-Entity Level** ☐ **Patient or Encounter-Level** ☒ **Both**

13. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements? NOTE** that data element validation from the literature is acceptable.

    **Submission document*:*** Questions 2b.01-02

    ☐ **Yes**

    ☒ **No**

    ☐ **Not applicable** (patient/encounter level testing was not performed)

14. **Method of establishing validity at the accountable-entity level:**

    **NOTE:** Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

    **Submission document:** Questions 2b.01-02

    ☒ **Face validity**

    ☒ **Empirical validity testing at the accountable-entity level**

    ☐ **N/A (accountable-entity level testing not conducted)**

15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    **Submission document:** Question 2b.02

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (accountable-entity level testing was not performed)

16. **Assess the method(s) for establishing validity**

    **Submission document**: Question 2b.02

- The developer included multiple approaches to establishing validity: face validity, patient-level validity testing, and empirical validity testing of the measure score.

Patient/Encounter-Level Validity

- During the early years of measure use (2017-2019), the developer worked with participating hospitals to audit all cases of inappropriate diagnoses (N=3197).
- Using the current measure as specified and data from 29 hospitals from 2020-2021, the senior HMS project manager audited 50 consecutive cases of patients diagnosed with UTI (appropriate or inappropriate). They calculated the proportion of data elements that had been abstracted correctly. Inter-rater reliability was used to assess the different between the original case classification and the audit ("correct") case classification, with data errors identified.
- Using the current measure as specified, 2-4 physicians conducted a structured implicit case review of inappropriate diagnosis of UTI using 2020 data. Cases were randomly selected from among "gray areas" (i.e. less certain categorizations identified during initial measure development, such as patients with altered mental status). Reviewers independently assessed their level of agreement with the classification of inappropriate diagnoses of UTI. Developer states, "If there was disagreement in classification, a discussion would commence that included ways to improve the measure to account for any errors in classification."

Face Validity

- Face validity was assessed again in 2021 using the current measure, receiving feedback from the 40 HMS hospitals used in the testing data, an 11-person technical expert panel (TEP), who include Infectious Diseases physicians, pharmacists, urologists, 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, and with a focus group of seven patients or caregivers.
- Face validity conducted with patients: The developer concluded from focus group discussions with patients and caregivers that this group understood the meaning of overdiagnosis and felt that measuring inappropriate diagnosis of UTI was both important and meaningful.

Empirical Validity of Measure Score

- The developer tested whether better performance on this score was related to better performance on a measure of inappropriate diagnosis of community-acquired pneumonia (CAP). No suitable structural or outcome measures were found for this correlation so using a literature review and consulting with experts, the developer determined these two measure both represent signals of hospital quality that affect patient outcomes.
- The developer also assessed the association of inappropriate diagnosis with antibiotic-associated adverse events.
- The developer next compared outcomes in patients inappropriately diagnosed with UTI vs. those who had ASB but were not treated with antibiotics using logistic generalized estimating equation models, inverse probability of treatment weighted by baseline covariates identified as significant, and other factors potentially associated with the outcome. Outcomes assessed included 30-day mortality, 30-day hospital readmission, 30-day emergency department visit, discharge to post-acute settings, Clostridioides difficile (C-Diff) infection at 30 days, and duration of hospitalization after urine testing.

17. **Assess the results(s) for establishing validity**

**Submission document:** Questions 2b.03-04

Face Validity

- Face validity conducted during measure development was used to help refine the measure and inform its specifications.
- Face validity conducted in 2021 was as follows:
    - Local leaders or quality champions from all 40 participating hospitals responded to the question, "Approximately, what percentage of cases called ASB by HMS do you agree are inappropriately diagnosed with ASB (0-100%)." The median response was 90 percent, inter-quartile range was 80 – 97 percent.
    - The TEP was asked to respond to the following statement, "The inappropriate diagnosis of UTI measure as specified can be used to distinguish between better and worse quality hospitals," using a 5-point scale Likert scale: 1=Strongly disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly agree.
        - Nine respondents agreed or strongly agreed with the statement.
        - One respondent was neutral.
        - One respondent strongly disagreed with the statement.
    - The TEP was also asked, "What additional data would you like to see captured related to the inappropriate diagnoses of UTI?"
        - Eight respondents said no additional data was needed.
        - One respondent would like to know the duration of antibiotic treatment. (The developer added this information to the measure submission in response.)
        - One respondent would like to see a balancing measure. (The developer added resources for studies on underdiagnosis of UTI to measure submission.)
        - One respondent would like to see length-of-stay data. (The developer added this information to the measure submission in response.)
    - Face validity conducted with patients: The developer concluded from focus group discussions with patients and caregivers that this group understood the meaning of overdiagnosis and felt that measuring inappropriate diagnosis of UTI was both important and meaningful.

Patient/Encounter-Level Validity

- Using 2020 data, 2-4 physicians performed a structured implicit case review of 25 cases of inappropriate diagnosis of UTI. The K for reviewer agreement was 0.72, which the developer states indicates substantial agreement.
    - 23/25 cases had 100% agreement by reviewers that the cases represented inappropriate diagnosis.
    - The developer notes that since this review used only "gray area" cases rather than a random selection, the true K may be even higher.
    - Two additional group of patients were added to the exclusion criteria as a result of this testing: those who were never treated for a UTI even if symptomatic, and those who only received antibiotics outside of the collection window. Also, "hypogastric" was added as a synonym for "suprapubic" in the UTI symptom listing.

Empirical Validity of the Measure Score

- 24,483 patients from 46 hospitals were analyzed for this analysis, using data from 7/1/17-3/31/20. The developer found that inappropriate diagnoses of UTI has a moderate correlation with inappropriate diagnosis of CAP at the hospital level (R=0.53, P<0.001).
- The findings were similar for the 2,049 patients initially inappropriately diagnosed within the emergency room setting (R=0.46, p=0.002).
- When examining the association of inappropriate diagnosis of UTI with adverse outcomes, the developer found that in 2,733 hospitalized patients inappropriately diagnosed with UTI were treated with a median seven days of antibiotic therapy (IQR: 4-9 days), all of which were unnecessary.
    - The median length of stay for patients inappropriately diagnosed with UTI was 4 days and for those not treated with antibiotics, 3 days. The adjusted odds ratio for those treated with antibiotics was 1.37 (95% Confidence Interval (C.I.) 1.28-1.47, p<0.001).
    - Other associations showed potentially increased odds for those who received antibiotics (e.g. postdischarge mortality, postdischarge readmission, and discharge to post-acute care facility) but these associations were not statistically significant.

### VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**

**Submission document:** Questions 2b.15-18

- The developer was able to estimate the percent of exclusions from the literature and directly from the HMS dataset. Total exclusions were estimated at 1.8%-2.47%.
- Exclusion criteria was not tested. The criteria were provided to participating hospitals and the TEP to ensure they appeared feasible and reasonable. Some TEP members suggested additional populations to include in the future but believed that starting with a less contentious group (i.e., hospitalized medical patients) would be a good start before moving into more difficult populations (e.g., nursing homes).

19. **Risk Adjustment**

**Submission Document:** Questions 2b.19-32

19a. **Risk-adjustment method**

☒ None ☐ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☐ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

19d.**Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)

☐ Yes ☐ No

19d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

19e**. Assess the risk-adjustment approach**

20. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    **Submission document:** Questions 2b.05-07

    *For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?*

    - Hospitals in the top 10th percentile for performance have approximately 12 fewer patients inappropriately diagnosed with UTI per 100 patients tread for UTI than the median. Those in the lowest 90th percentile for performance have about 15 more patients inappropriately diagnosed per 100 patients treated for UTI than the median. Hospitals were also analyzed by quartiles and the developer found a statistically significant difference in performance between all adjacent quartiles.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    **Submission document:** Questions 2b.11-14

    - N/A

22. **Please describe any concerns you have regarding missing data.**

    **Submission document:** Questions 2b.08-10

    - The developer states that the percentage of encounters with missing data (labeled "unknown" or "not available" in the abstraction tool) was 5.2 percent (714/13,805). The majority of missing demographic data was for the ethnicity category, missing 14.6 percent (2,019/13,805). The majority of missing data relevant to UTI was for urinary catheter, missing 3.5 percent (484/13,805). The developer states that most missing data did not exist in the medical record and was not due to errors in abstraction.

**For cost/resource use measures ONLY:**

*If not cost/resource use measure, please skip to question 25*

23. **Are the specifications in alignment with the stated measure intent?**

    **☐ Yes ☐ Somewhat ☐ No (If "Somewhat" or "No", please explain)**

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats. Adj-noun noun**

    ☐ **High** (NOTE: Can be HIGH only if accountable entity level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable entity level testing has NOT been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Potential threats to validity are empirically assessed (Box 1) → Empirical validity testing was conducted using the measure as specified (Box 2) → Empirical validity testing conducted at the accountable entity level (Box 4) → Testing method appropriate (Box 6) → Based on empirical testing there is moderate confidence the measure is valid (Box 7b) → Rate as MODERATE

### FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

**Submission documents:** Questions 2c.01-08

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

### ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

## Criteria 1: Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria

---

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

***2021 Submission:***

Updated evidence information here.

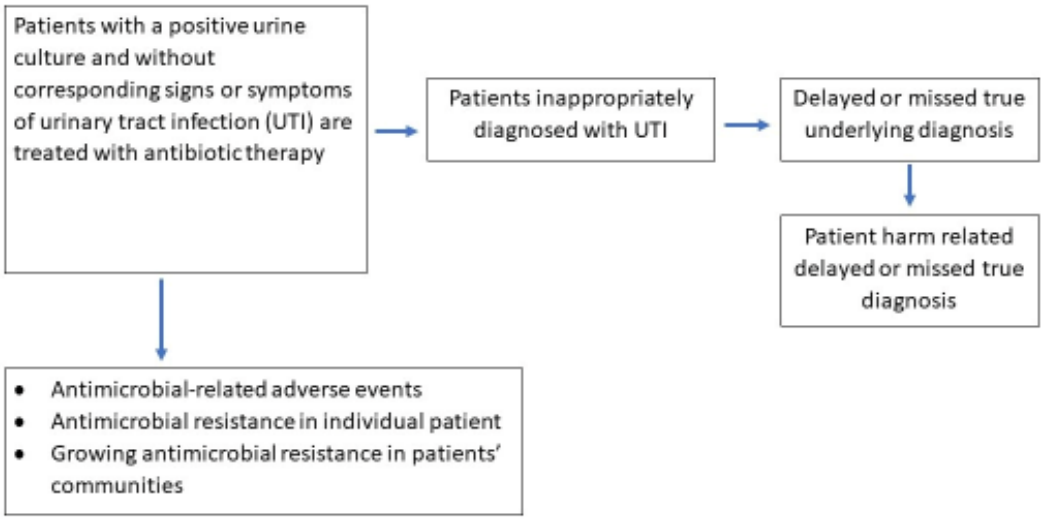
***2018 Submission:***

Evidence from the previous submission here.

**1a.01. Provide a logic model.**

*Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.*

**[Response Begins]**




**[Response Ends]**

**1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.**

*A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data.*

**[Response Begins]**

Clinical Practice Guideline recommendation (with evidence review)

**[Response Ends]**

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

**Evidence - Systematic Reviews Table (Repeatable)**

Group 1 - Evidence - Systematic Reviews Table

**1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.**

**[Response Begins]**

Clinical Practice Guideline for the Management of Asymptomatic Bacteriuria: 2019 Update by the Infectious Diseases Society of America. Infectious Disease Society of America. 15 May 2019.

Nicolle LE, Gupta K, Bradley SF, Colgan R, DeMuri GP, Drekonja D, Eckert LO, Geerlings SE, Köves B, Hooton TM, Juthani-Mehta M, Knight SL, Saint S, Schaeffer AJ, Trautner B, Wullt B, Siemieniuk R. Clinical Practice Guideline for the Management of Asymptomatic Bacteriuria: 2019 Update by the Infectious Diseases Society of America. Clin Infect Dis. 2019 May 2;68(10):e83-e110. doi: 10.1093/cid/ciy1121. PMID: 30895288.

https://www.idsociety.org/practice-guideline/asymptomatic-bacteriuria/

**[Response Ends]**

**1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.**

**[Response Begins]**

**II. Should ASB (asymptomatic bacteriuria) be screened for or treated in healthy nonpregnant women?**

1. In healthy premenopausal, nonpregnant women or healthy postmenopausal women, we recommend against screening for or treating ASB *(strong recommendation, moderate-quality evidence)*.

**IV. Should ASB Be Screened for and Treated in Functionally Impaired Older Women or Men Residing in the Community, or in Older Residents of Longterm Care Facilities? Recommendations**

1. In older, community-dwelling persons who are functionally impaired, we recommend against screening for or treating ASB (*strong recommendation, low-quality evidence*).
2. In older persons resident in long-term care facilities, we recommend against screening for or treating ASB (*strong recommendation, moderate-quality evidence*)

**V. In an older, functionally or cognitively impaired patient, which nonlocalizing symptoms distinguish ASB from symptomatic UTI?**

1. In older patients with functional and/or cognitive impairment with bacteriuria and delirium (acute mental status change, confusion) and without local genitourinary symptoms or other systemic signs of infection (e.g., fever or hemodynamic instability), we recommend assessment for other causes and careful observation rather than antimicrobial treatment *(strong recommendation, very low-quality evidence)*.

**VI. Should patients with diabetes be screened or treated for ASB?**

1. In patients with diabetes, we recommend against screening for or treating ASB *(strong recommendation, moderate-quality evidence)*.

**XI. Should patients with an indwelling urethral catheter be screened or treated for ASB?**

1. In patients with a short-term indwelling urethral catheter (<30 days), we recommend against screening for or treating ASB *(strong recommendation, low-quality evidence)*. **Remarks:** Considerations are likely to be similar for patients with indwelling suprapubic catheters, and it is reasonable to manage these patients similar to patients with indwelling urethral catheters, for both short-term and long-term suprapubic catheterization.
2. In patients with long-term indwelling catheters, we recommend against screening for or treating ASB *(strong recommendation, low-quality evidence)*.

**[Response Ends]**

**1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.**

**[Response Begins]**

Grading of Recommendations Assessment, Development and Education (GRADE) approach for rating the confidence and the evidence.

**Table 1**. Five factors that influence confidence in evidence

| Factor | Description |
|---|---|
| Study limitations | Severity of threats to studies' internal validity (e.g., randomized vs observational design, potential for confounding, bias in measurement) |
| Inconsistency of results | Do different studies provide similar or different estimates of effect size |
| Indirectness of evidence | How relevant are the studies to the clinical question at hand (e.g., nature of study population, comparison group, type of outcomes measured) |
| Imprecision | Precision of estimates of effect |
| Reporting bias | Risk of bias due to selective publication of results |

**Table 2**. GRADE quality of evidence definitions (can be modified by confidence factors, above)

| Quality of Evidence | Definition |
|---|---|
| High Quality | Further research is very unlikely to change our confidence in the estimate of effect |
| Moderate Quality | Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate |
| Low Quality | Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate |
| Very Low Quality | Any estimate of effect is very uncertain |

1. Guyatt GH, Oxman AD, Vist GE, et al. ; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336:924–6.

2.Jaeschke R, Guyatt GH, Dellinger P, et al. ; GRADE Working Group. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ 2008; 337:a744.

3.Schünemann HJ, Oxman AD, Brozek J, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. Evid Based Med 2008; 13:162–3.

The grades assigned by Infectious Disease Society of America clinical practice guidelines to the quality of evidence varied by recommendation and are summarized in Table 3, below.

**Table 3.** Quality of evidence grades assigned by Infectious Disease Society of America (IDSA) clinical practice guidelines.

| Recommendation | Evidence Grade |
|---|---|
| II.1. In healthy premenopausal, nonpregnant women or healthy postmenopausal women, we recommend against screening for or treating ASB | Moderate-quality evidence, defined as "Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate" |
| IV.1. In older, community-dwelling persons who are functionally impaired, we recommend against screening for or treating ASB | Low-quality evidence, defined as "Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate" |
| IV.2. In older persons resident in long-term care facilities, we recommend against screening for or treating ASB | Moderate-quality evidence, defined as "Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate" |
| V.1. In older patients with functional and/or cognitive impairment with bacteriuria and delirium (acute mental status change, confusion) and without local genitourinary symptoms or other systemic signs of infection (e.g., fever or hemodynamic instability), we recommend assessment for other causes and careful observation rather than antimicrobial treatment | Very low-quality evidence, defined as "Any estimate of effect is very uncertain" |
| VI.1. In patients with diabetes, we recommend against screening for or treating ASB | Moderate-quality evidence, defined as "Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate" |
| XI.1. In patients with a short-term indwelling urethral catheter (<30 days), we recommend against screening for or treating ASB **Remarks:** Considerations are likely to be similar for patients with indwelling suprapubic catheters, and it is reasonable to manage these patients similar to patients with indwelling urethral catheters, for both short-term and long-term suprapubic catheterization | Low-quality evidence, defined as "Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate" |
| X1.3. In patients with long-term indwelling catheters, we recommend against screening for or treating ASB | Low-quality evidence, defined as "Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate" |

**[Response Ends]**

**1a.06. Provide all other grades and definitions from the evidence grading system.**

**[Response Begins]**

High Quality Evidence - Further research is very unlikely to change our confidence in the estimate of effect

Figure 1. Quality of evidence was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE).

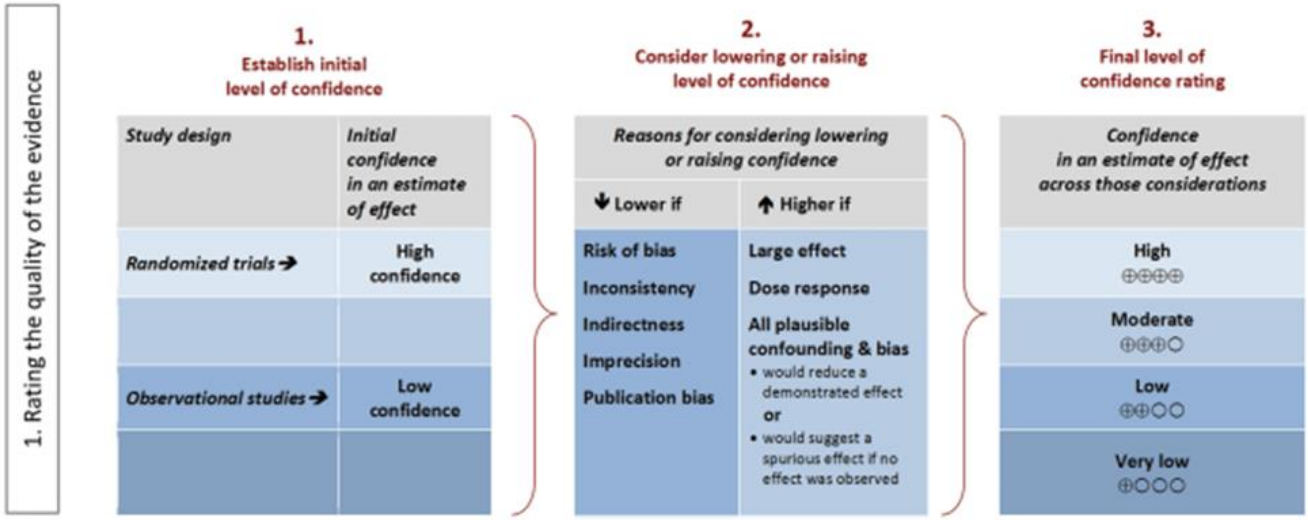



Figure from the US GRADE Network

**[Response Ends]**

**1a.07. Provide the grade assigned to the recommendation, with definition of the grade.**

**[Response Begins]**

**Table 4.** Grade assigned to recommendations by Infectious Disease Society of America (IDSA) clinical practice guidelines.

| Recommendation | Recommendation Grade |
|---|---|
| II.1. In healthy premenopausal, nonpregnant women or healthy postmenopausal women, we recommend against screening for or treating ASB | Strong recommendation - Moderate- or high-quality evidence that the desirable consequences outweigh the undesirable consequences for a course of action.<br>"There is high-quality evidence that antibiotics have an increased risk of adverse effects, that screening and treating ASB is extremely costly, and that the use of antibiotics promotes emergence of antimicrobial resistance." |

| Recommendation | Recommendation Grade |
|---|---|
| IV.1. In older, community-dwelling persons who are functionally impaired, we recommend against screening for or treating ASB | Strong recommendation - Moderate- or high-quality evidence that the desirable consequences outweigh the undesirable consequences for a course of action.<br>"We make strong recommendations because there is low- or moderate-quality evidence that there is no benefit and high-quality evidence of harm." |
| IV.2. In older persons residing in long-term care facilities, we recommend against screening for or treating ASB | Strong recommendation - Moderate- or high-quality evidence that the desirable consequences outweigh the undesirable consequences for a course of action.<br>"We make strong recommendations because there is low- or moderate-quality evidence that there is no benefit and high-quality evidence of harm." |
| V.1. In older patients with functional and/or cognitive impairment with bacteriuria and delirium (acute mental status change, confusion) and without local genitourinary symptoms or other systemic signs of infection (e.g., fever or hemodynamic instability), we recommend assessment for other causes and careful observation rather than antimicrobial treatment | Strong recommendation - Moderate- or high-quality evidence that the desirable consequences outweigh the undesirable consequences for a course of action.<br>"We make a strong recommendation because there is high certainty for harm and low certainty of any benefit from treatment of ASB in older adults." |
| VI.1. In patients with diabetes, we recommend against screening for or treating ASB | Strong recommendation - Moderate- or high-quality evidence that the desirable consequences outweigh the undesirable consequences for a course of action.<br>"Based on the lack of demonstrated benefit and the possible harms that occur with additional antimicrobial use, we recommend against screening for or treating ASB in persons with diabetes." |
| XI.1. In patients with a short-term indwelling urethral catheter (<30 days), we recommend against screening for or treating ASB **Remarks:** Considerations are likely to be similar for patients with indwelling suprapubic catheters, and it is reasonable to manage these patients similar to patients with indwelling urethral catheters, for both short-term and long-term suprapubic catheterization | Strong recommendation - Moderate- or high-quality evidence that the desirable consequences outweigh the undesirable consequences for a course of action.<br>Whether or not antimicrobials for ASB are effective in preventing symptomatic UTI, sepsis, or death is uncertain. In the acute care hospital setting, the risk of *Clostridioides difficile* infection is high; thus, avoiding antimicrobials is particularly important in hospitalized patients. |

| Recommendation | Recommendation Grade |
|---|---|
| X1.3. In patients with long-term indwelling catheters, we recommend against screening for or treating ASB | Strong recommendation - Moderate- or high-quality evidence that the desirable consequences outweigh the undesirable consequences for a course of action.<br>Whether there is a benefit of antimicrobial therapy for ASB while a catheter remains in situ is uncertain (low-quality evidence), and there is high-quality evidence of harm with increased antimicrobial resistance. |

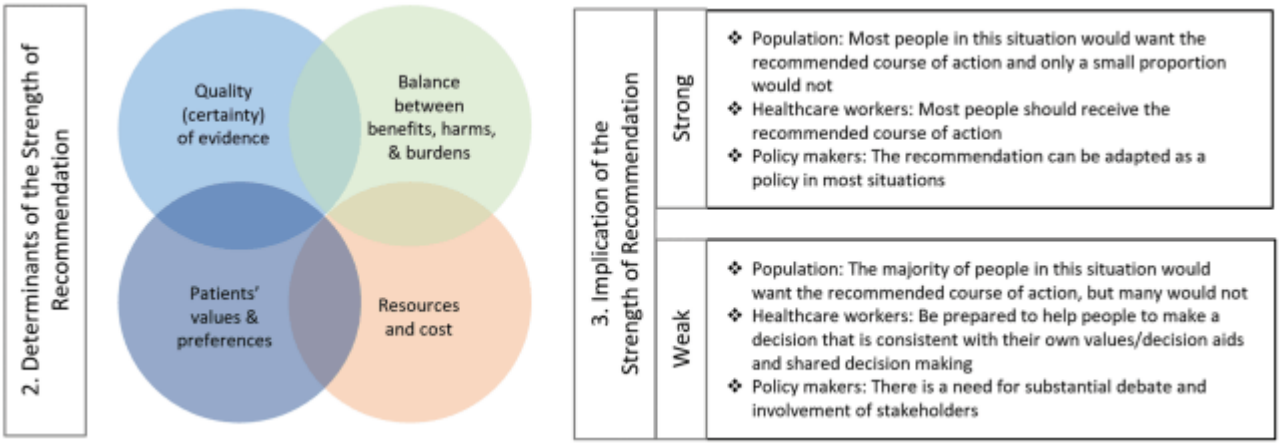
**[Response Ends]**

**1a.08. Provide all other grades and definitions from the recommendation grading system.**

**[Response Begins]**

All recommendations followed GRADE and are either "strong" or "weak". Figure 2 outlines how strength of recommendations was determined and implications of the strength of recommendations.

**Figure 2.** GRADE determinants of the strength of recommendations and implications.




A recommendation was graded as "Strong" when:

- Moderate- or high-quality evidence that the desirable consequences outweigh the undesirable consequences for a course of action, OR
- High-quality evidence of harm and benefits are uncertain (i.e., low or very low quality)

A recommendation was graded as "Weak" when conditions for strong recommendation were not met.

**Table 5**. Interpretation of Strong and Weak (Conditional) Recommendations by Patients, Clinicians, and Policy Makers

| Stakeholder | Strong Recommendation | Weak (Conditional) Recommendation |
|---|---|---|
| Patients | All or almost all individuals in this situation would want the recommended course of action, and only a small proportion would not. | Most individuals in this situation would probably want the suggested course of action, but many would not. |

| Stakeholder | Strong Recommendation | Weak (Conditional) Recommendation |
|---|---|---|
| Clinicians | All or almost all individuals should receive the intervention. | Recognize that fully informed individuals might reasonably choose different courses of action. A shared decision-making process is typically useful in helping individuals to make decisions consistent with their values and preferences. |
| Policy makers | The recommendation can be adopted as policy in most situations. Adherence to this recommendation according to the guideline can be used as a quality criterion or performance indicator. | Policymaking will require substantial debate and involvement of various stakeholders. |

**[Response Ends]**

**1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.**

**[Response Begins]**

**Table 6**. Quantity and type of studies in support of each recommendation.

| Recommendation | Number of Studies | Study Type (N) |
|---|---|---|
| II.1. In healthy premenopausal, nonpregnant women or healthy postmenopausal women, we recommend against screening for or treating ASB | 7 | Randomized clinical trial (2)<br>Retrospective cohort (2)<br>Conference report (1)<br>Clinical Practice Guideline (1)<br>Systematic Review (1) |
| IV.1. In older, community-dwelling persons who are functionally impaired, we recommend against screening for or treating ASB | 5 | Randomized clinical trial (4)<br>Cohort study and clinical trial (1) |
| IV.2. In older persons resident in long-term care facilities, we recommend against screening for or treating ASB | 14 | Prospective cohort study (5)<br>Comparative study (3)<br>Clinical practice guideline (2)<br>Clinical trial (1)<br>Consensus conference report (1)<br>Cross-sectional study (1)<br>Retrospective cohort (1) |

| Recommendation | Number of Studies | Study Type (N) |
|---|---|---|
| V.1. In older patients with functional and/or cognitive impairment with bacteriuria and delirium (acute mental status change, confusion) and without local genitourinary symptoms or other systemic signs of infection (e.g., fever or hemodynamic instability), we recommend assessment for other causes and careful observation rather than antimicrobial treatment | 14 | Prospective cohort study (4)<br>Clinical practice guideline (2)<br>Cross-sectional study (2)<br>Retrospective cohort (2)<br>Systematic review (2)<br>Consensus conference report (1)<br>Randomized clinical trial (1) |
| VI.1. In patients with diabetes, we recommend against screening for or treating ASB | 5 | Prospective cohort study (3)<br>Clinical practice guideline (1)<br>Randomized clinical trial (1) |
| XI.1. In patients with a short-term indwelling urethral catheter (<30 days), we recommend against screening for or treating ASB **Remarks:** Considerations are likely to be similar for patients with indwelling suprapubic catheters, and it is reasonable to manage these patients similar to patients with indwelling urethral catheters, for both short-term and long-term suprapubic catheterization | 10 | Prospective cohort study (3)<br>Randomized clinical trial (2)<br>Case-control study (1)<br>Cochrane review (1)<br>Clinical practice guideline (1)<br>Retrospective cohort study (1)<br>Systematic review (1) |
| X1.3. In patients with long-term indwelling catheters, we recommend against screening for or treating ASB | 9 | Prospective cohort study (5)<br>Retrospective cohort study (2)<br>Randomized clinical trial (1)<br>Randomized intervention trial, QI (1) |

**[Response Ends]**

**1a.10. Provide the estimates of benefit, and consistency across studies.**

**[Response Begins]**

This publication did not provide an estimate of benefit or consistency across the cited studies.

**[Response Ends]**

**1a.11. Indicate what, if any, harms were identified in the study.**

**[Response Begins]**

Within the clinical practice guideline, authors considered values and preferences from the viewpoint of the patient and the societal perspective. While most patients may wish to receive antimicrobial therapy for ASB where benefits outweigh

harms, they highlight the significant potential harms related to antimicrobial therapy, including adverse drug effects, *Clostridioides difficile* infection, and the potential for antimicrobial resistance. They state that while the evidence quality is generally low and there is no suggestion of potential harm, the general recommendation for not treating ASB stems from high-quality evidence that antimicrobial therapy contributes to antimicrobial resistance.

**[Response Ends]**

**1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.**

**[Response Begins]**

Since publication of the guidelines, there have been additional studies evaluating trends in inpatient use[1] and multi-drug resistant bacterial infections in US hospitalized patients.[2] For example, one study suggests that between 2012 and 2017, overall antibiotic days of therapy in US hospitals were unchanged.[1] While the prevalence of some multi-drug resistant bacteria decreased over that time period (e.g., methicillin-resistant staphylococcus aureus (MRSA)), other highly concerning multi-drug resistant organisms flourished. For example, the incidence of infections resulting from extended-spectrum beta-lactamase (ESBL) producing organisms increased by 53.3% (from 37.55 to 57.12 cases per 10,000 hospitalizations).[2] Another recent study found that recent antibiotic exposure was positively associated with baseline multi-drug resistant colonization (odds ratio 1.70; 95% confidence interval 1.22-2.38).[3] These studies add further support to the IDSA guidelines referenced above.

[1] James Baggs, PhD, Sophia Kazakova, MD, MPH, PhD, Kelly M Hatfield, MSPH, Sujan Reddy, MD, MSc, Arjun Srinivasan, MD, Lauri Hicks, DO, Melinda M Neuhauser, PharmD, MPH, John A Jernigan, MD, MS, 2891. Trends in Inpatient Antibiotic Use in US Hospitals, 2012–2017, *Open Forum Infectious Diseases*, Volume 6, Issue Supplement_2, October 2019, Page S79,

[2]Jernigan JA, Hatfield KM, Wolford H, Nelson RE, Olubajo B, Reddy SC, McCarthy N, Paul P, McDonald LC, Kallen A, Fiore A, Craig M, Baggs J. Multidrug-Resistant Bacterial Infections in U.S. Hospitalized Patients, 2012-2017. N Engl J Med. 2020 Apr 2;382(14):1309-1319.

[3] Gontjes KJ, Gibson KE, Lansing BJ, Mantey J, Jones KM, Cassone M, Wang J, Mills JP, Mody L, Patel PK. Association of Exposure to High-risk Antibiotics in Acute Care Hospitals With Multidrug-Resistant Organism Burden in Nursing Homes. JAMA Netw Open. 2022 Feb 1;5(2):e2144959.

**[Response Ends]**

**1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.**

**[Response Begins]**

Our definition of inappropriate diagnosis of UTI is based on treatment for urinary tract infection in the absence of meeting criteria for UTI. The criteria for urinary tract infection within our measure is consistent with the National Healthcare Safety Network (NHSN) classification of UTI.[1]

Table 7. Comparison of NHSN classification of UTI to the inappropriate diagnosis of UTI measure

| Condition | NHSN Definition | Changes associated with submitted measure |
|---|---|---|
| Non-catheter-associated UTI in any age patient | Patient must meet 1, 2, and 3 below:<br>1. One of the following is true:<br>• Patient has/had an indwelling urinary catheter, but it has/had not been in place for more than two consecutive days in in inpatient location on the date of the event, OR<br>• Patient did not have an indwelling urinary catheter in place on the date of the event nor the day before the date of the event<br>2. Patient has at least one of the following signs or symptoms:<br>• Fever (>38 degrees Celsius)<br>• Suprapubic tenderness with no other recognized cause<br>• Costovertebral angle pain or tenderness with no other recognized cause<br>• Urinary frequency<br>• Urinary urgency<br>• Dysuria<br>3. Patient has a urine culture with no more than two species of organisms identified, at least one of which is a bacterium of $\geq 10^5$ CFU/ml. | No differences |

| Condition | NHSN Definition | Changes associated with submitted measure |
|---|---|---|
| Catheter-associated Urinary Tract Infection (CAUTI) in any age patient | Patient 1, 2, and 3 below:<br>1. Patient had an indwelling urinary catheter that had been in place for more than 2 consecutive days in an inpatient location on the date of the event AND was either:<br>• Present for any proportion of the calendar day on the date of the event, OR<br>• Removed the day before the date of the event<br>2. Patient has at least one of the following signs or symptoms:<br>• Fever (>38 degrees Celsius)<br>• Suprapubic tenderness with no other recognized cause<br>• Costovertebral angle pain or tenderness with no other recognized cause<br>• Urinary frequency<br>• Urinary urgency<br>• Dysuria<br>3. Patient has a urine culture with no more than two species of organisms identified, at least one of which is a bacterium of $\geq 10^5$ CFU/ml. | While the NHSN guidelines are designed to facilitate identifying hospital-acquired infections, our measure targets patients either at the time of admission or during their hospitalization. As a result, we do not require that a urinary catheter be in place for 2 consecutive days in an inpatient location, rather our measure allows for catheter presence in either the inpatient or outpatient location.<br>We also have a shorter "window period" in which to evaluate for symptoms. We discussed with our Technical Expert Panel whether to be consistent with the NHSN (-3 to +3 days) or to narrow the window to -1 to + 2 days (with day 0 being the day of urine culture collection). The data using both window periods were quite similar but the time needed for data collection (and thus feasibility) would be reduced if we moved to a -1 to +2 range. All experts agreed with reducing to -1 to +2, noting that NHSN was focused on healthcare-associated infections whereas these inappropriate diagnoses are often community-associated and therefore a longer range for symptoms collection is not typically needed. |

Similarly, the Clinical Practice Guideline for the Management of Asymptomatic Bacteriuria: 2019 Update by the Infectious Diseases Society of America defines asymptomatic bacteriuria as "the presence of 1 or more species of bacteria growing in the urine at specified quantitative counts ($\geq 10^5$ colony-forming units [CFU]/mL or $\geq 10^8$ CFU/L), irrespective of the presence of pyuria, in the absence of signs or symptoms attributable to urinary tract infection (UTI). This definition of ASB, or absence of UTI, is consistent with the submitted measure.[2] These clinical practice guidelines, discussed in more detail in 1a.03-1a.12, highlight the lack of benefit of treatment of ASB in several populations, described further in 1a.14. We also use their criteria of when to treat altered mental status as a UTI: 1) when altered mental status occurs with other symptoms or 2) when patient has "other systemic signs of infection (e.g., fever or hemodynamic instability)."[2] 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]

**[Response Ends]**

**1a.14. Briefly synthesize the evidence that supports the measure.**

**[Response Begins]**

***Antibiotics and ASB***

Lack of benefit for treatment of ASB has been demonstrated in several populations. In a study of 673 consecutively enrolled, asymptomatic women, aged 18-40, from January 2005 to December 2009, ASB recurrence (based on microbiological and clinical information) was significantly higher in the group treated for ASB at 6 months (relative risk [RR] 1.31; 95% CI, 1.21-1.42, P<0.0001), and at 12 months (RR 3.17; 95% CI, 2.55-3.90; p<0.0001).[4] In a study of women (> 16 years of age) with diabetes, bacteriuria, and no symptoms, patients treated for ASB had lower rates of bacteriuria at four weeks, but no difference in frequency of symptomatic urinary tract infection, time to first urinary tract infection, pyelonephritis, or hospitalization for urinary tract infection during a mean follow-up period of 27 months. Patients in the antimicrobial-therapy group had nearly five times as many days of antibiotic use as those in the placebo group.[5] In another study of patients aged ≥18 years who had an indwelling urinary catheter in place for at least one week, patients with cephalexin-sensitive bacteriuria receiving cephalexin had similar rates of fever but higher rates of growth of cephalexin-resistant urinary bacteria than those not receiving antibiotics.[6] An additional study of elderly institutionalized women (mean age 83.4 years +/- 8.8 years) with asymptomatic bacteriuria showed that those assigned to receive antibiotic therapy for all episodes of bacteriuria has higher rates incidence of reinfection (1.67 versus 0.87 per patient year) and adverse antimicrobial drug effects (0.51 versus 0.046 per patient year) compared to those only receiving therapy with development of symptoms.[7] Another study evaluating hospitalized adult patients treated for ASB noted no differences in 30-day post-discharge mortality, readmissions, or ED visits, though did note a longer duration of hospitalization for patients treated for ASB than those not treated (median 4 vs 3 days, adjusted relative risk 1.37 [1.28-1.47], p<0.001).[8] Finally, while not directly related to treatment of ASB, the United States Preventative Services Taskforce recommends that men and nonpregnant women not be screened for ASB, providing a D grade, indicating that "the USPSTF recommends against this service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits."[9]

***Delayed Time to True Diagnosis***

Diagnostic errors are associated with poor outcomes including longer length of stay, higher inpatient mortality, increased ICU admission, and higher 30-day readmission. Diagnostic errors are also the most common cause of malpractice claims, are financially costly, and—compared to other claims—result in the most harm. Specifically, diagnostic error related to infections account for 13.5% of high-severity diagnostic error cases (i.e., resulting in serious, permanent disability or death).[10-15] Specifically, we found that inappropriate diagnosis of UTI is associated with a 1 day longer length of stay after urine testing.[8] This delay is likely the result of waiting for urine culture sensitivities to return. Each additional day of hospitalization costs approximately $2,607 making this harm very costly to hospitals and health systems. If you estimate 400,000 hospitalizations for UTI every year[16] of which 1/3 inappropriately diagnosed[3] and result in a 1 day increase in LOS[8] the annual excess cost in the US alone of inappropriate diagnosis of UTI is approximately $326 million (estimated cost $2,607/hospital day). Another prospective study actually found a trend toward higher 30-day readmission (9.6% vs. 6.3%) and mortality (7.0% vs. 4.8%) in patients inappropriately diagnosed with UTI when compared to patients with ASB who were not treated with antibiotics.[17]

***Antibiotic Associated Adverse Events***

In our HMS cohort, we found that up to 3% of patients reported antibiotic-associated adverse events after being inappropriately diagnosed with UTI. Deep chart review by trained clinicians suggests that number is far higher and that up to 20% of hospitalized patients treated unnecessarily with antibiotics develop an antibiotic-associated adverse-event.[18] Most common are gastrointestinal, renal, and hematologic abnormalities, accounting for 42%, 24%, and 15% of 30-day ADEs respectively. In addition, 1-3% of patients treated with antibiotics in the hospital setting will develop *Clostridioides difficile* infections. One prospective study found that patients inappropriately diagnosed with UTI had a 30-day *Clostridioides difficile* infection rate of 1.3% vs. those with AS not treated with antibiotics.[19] Fortunately, many of these harms could be reduced with better antibiotic use. Improving appropriate antibiotic use for UTI can reduce patient-reported antibiotic-associated adverse events and ecological or retrospective studies have suggested that up to 60% of *Clostridioides difficile* infections (and related mortality) could be eliminated by improving antibiotic prescribing.[20]

***Antibiotic Use and Antimicrobial Resistance***

Antibiotic use in patients inappropriately diagnosed with infections continues to be a large driver of antibiotic use and antibiotic resistance. Between 2012 and 2017, overall antibiotic days of therapy in US hospitals were unchanged.[21] While

the prevalence of some multi-drug resistant bacteria decreased over that time period (e.g., methicillin-resistant staphylococcus aureus (MRSA)), other highly concerning multi-drug resistant organisms flourished. For example, the incidence of infections resulting from extended-spectrum beta-lactamase (ESBL) producing organisms increased by 53.3% (from 37.55 to 57.12 cases per 10,000 hospitalizations).[22] A systematic review and meta-analysis of the literature found a significant positive relationship between antibiotic consumption and development of antimicrobial resistance, with a pooled odds ratio of 2.3 (95% confidence interval 2.2-2.5).[23] Similarly, a recent study found that recent antibiotic exposure was positively associated with baseline multi-drug resistant colonization (odds ratio 1.70; 95% confidence interval 1.22-2.38).[24] The implications of these antimicrobial resistant organisms are significant. Globally, predictive statistical models estimate 4.95 million (3.62-6.57 million) deaths associated with bacterial antimicrobial resistance in 2019, of which 1.27 million (95% uncertainly interval 0.911-1.71) deaths were attributable to bacterial antimicrobial resistance.[25] In a study of inpatient admissions in the US Department of Veterans Affairs between October 2007 and November 2010, healthcare-associated infections (HAI) with multi-drug resistant gram negative bacteria were associated with a significantly elevated risk of mortality as was HAI or colonization with MRSA.[26]

**[Response Ends]**

**1a.15. Detail the process used to identify the evidence.**

**[Response Begins]**

Evidence was identified through appropriate clinical practice guidelines[2,9] and through comprehensive Pubmed search of studies as they pertain to diagnosis of ASB, UTI, or treatment of ASB.

**[Response Ends]**

**1a.16. Provide the citation(s) for the evidence.**

**[Response Begins]**

[1]Urinary Tract (Catheter-Associated Urinary Tract Infection [CAUTI] and Non-Catheter-Associated Urinary Tract Infections [UTI]) Events. National Healthcare Safety Network. Centers for Disease Control. January 2022. <https://www.cdc.gov/nhsn/pdfs/pscmanual/7psccauticurrent.pdf>

[2]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

[3] Nicolle LE, Gupta K, Bradley SF, Colgan R, DeMuri GP, Drekonja D, Eckert LO, Geerlings SE, Köves B, Hooton TM, Juthani-Mehta M, Knight SL, Saint S, Schaeffer AJ, Trautner B, Wullt B, Siemieniuk R. Clinical Practice Guideline for the Management of Asymptomatic Bacteriuria: 2019 Update by the Infectious Diseases Society of America. Clin Infect Dis. 2019 May 2;68(10):e83-e110. doi: 10.1093/cid/ciy1121. PMID: 30895288.

[4] Cai T, Mazzoli S, Mondaini N, Meacci F, Nesi G, D'Elia C, Malossini G, Boddi V, Bartoletti R. The role of asymptomatic bacteriuria in young women with recurrent urinary tract infections: to treat or not to treat? Clin Infect Dis. 2012 Sep;55(6):771-7.

[5] Harding GK, Zhanel GG, Nicolle LE, Cheang M; Manitoba Diabetes Urinary Tract Infection Study Group. Antimicrobial treatment in diabetic women with asymptomatic bacteriuria. N Engl J Med. 2002 Nov 14;347(20):1576-83.

[6] Warren JW, Anthony WC, Hoopes JM, Muncie HL. Cephalexin for Susceptible Bacteriuria in Afebrile, Long-term Catheterized Patients. *JAMA.* 1982;248(4):454–458

[7] Nicolle LE, Mayhew WJ, Bryan L. Prospective randomized comparison of therapy and no therapy for asymptomatic bacteriuria in institutionalized elderly women. Am J Med. 1987 Jul;83(1):27-33.

[8]Petty LA, Vaughn VM, Flanders SA, et al. Risk Factors and Outcomes Associated With Treatment of Asymptomatic Bacteriuria in Hospitalized Patients. *JAMA Intern Med*. 2019;179(11):1519-1527.

[9]Screening for Asymptomatic Bacteriuria in Adults: U.S. Preventive Services Tast Force Reaffirmation Recommendation Statement. U.S. Preventive Services Task Force, Agency for Healthcare Research and Quality, Rockville, Maryland. 1 July 2008.

[10] Eames J, Eisenman A, Schuster RJ. Disagreement between emergency department admission diagnosis and hospital discharge diagnosis: mortality and morbidity. Diagnosis (Berl). 2016;3(1):23-30. doi:10.1515/dx-2015-0028.

[11] Gupta A, Snyder A, Kachalia A, Flanders S, Saint S, Chopra V. Malpractice claims related to diagnostic errors in the hospital. BMJ Qual Saf. 2017;27(1). doi:10.1136/bmjqs-2017-006774.

[12] Johnson T, McNutt R, Odwazny R, Patel D, Baker S. Discrepancy between admission and discharge diagnoses as a predictor of hospital length of stay. Journal of Hospital Medicine. 2009;4(4):234-239. doi:10.1002/jhm.453.

[13] Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Serious misdiagnosis-related harms in malpractice claims: The "Big Three" - vascular events, infections, and cancers. Diagnosis (Berl). 2019;6(3):227-240. doi:10.1515/dx-2019-0019.

[14] Saber Tehrani AS, Lee H, Mathews SC, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National Practitioner Data Bank. BMJ Qual Saf. 2013;22(8):672-680. doi:10.1136/bmjqs-2012-001550.

[15] Winters B, Custer J, Galvagno SM, Jr., et al. Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. BMJ Qual Saf. 2012;21(11):894-902. doi:10.1136/bmjqs-2012-000803.

[16] Simmering JE, Tang F, Cavanaugh JE, Polgreen LA, Polgreen PM. The Increase in Hospitalizations for Urinary Tract Infections and the Associated Costs in the United States, 1998-2011. Open Forum Infect Dis. 2017;4(1):ofw281. doi:10.1093/ofid/ofw281. PCMID: PMC5414046.

[17] Spivak ES, Burk M, Zhang R, et al. Management of Bacteriuria in Veterans Affairs Hospitals. Clin Infect Dis. 2017;65(6):910-917. doi:10.1093/cid/cix474.

[18] Tamma PD, Avdic E, Li DX, Dzintars K, Cosgrove SE. Association of Adverse Events With Antibiotic Use in Hospitalized Patients. JAMA Intern Med. 2017.

[19] CDC. Nearly half a million Americans suffered from Clostridium difficile infections in a single year. U.S. Department of Health & Human Services; 02/25 2015.

[20] Thorpe KE, Joski P, Johnston KJ. Antibiotic-Resistant Infection Treatment Costs Have Doubled Since 2002, Now Exceeding $2 Billion Annually. Health Aff (Millwood). 2018;37(4):662-669. doi:10.1377/hlthaff.2017.1153.

[21] James Baggs, PhD, Sophia Kazakova, MD, MPH, PhD, Kelly M Hatfield, MSPH, Sujan Reddy, MD, MSc, Arjun Srinivasan, MD, Lauri Hicks, DO, Melinda M Neuhauser, PharmD, MPH, John A Jernigan, MD, MS, 2891. Trends in Inpatient Antibiotic Use in US Hospitals, 2012–2017, *Open Forum Infectious Diseases*, Volume 6, Issue Supplement_2, October 2019, Page S79,

[22] Jernigan JA, Hatfield KM, Wolford H, Nelson RE, Olubajo B, Reddy SC, McCarthy N, Paul P, McDonald LC, Kallen A, Fiore A, Craig M, Baggs J. Multidrug-Resistant Bacterial Infections in U.S. Hospitalized Patients, 2012-2017. N Engl J Med. 2020 Apr 2;382(14):1309-1319.

[23] Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. BMC Infect Dis. 2014 Jan 9;14:13.

[24] Gontjes KJ, Gibson KE, Lansing BJ, Mantey J, Jones KM, Cassone M, Wang J, Mills JP, Mody L, Patel PK. Association of Exposure to High-risk Antibiotics in Acute Care Hospitals With Multidrug-Resistant Organism Burden in Nursing Homes. JAMA Netw Open. 2022 Feb 1;5(2):e2144959.

[25]Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022 Feb 12;399(10325):629-655.

[26] Nelson RE, Slayton RB, Stevens VW, Jones MM, Khader K, Rubin MA, Jernigan JA, Samore MH. Attributable Mortality of Healthcare-Associated Infections Due to Multidrug-Resistant Gram-Negative Bacteria and Methicillin-Resistant Staphylococcus Aureus. Infect Control Hosp Epidemiol. 2017 Jul;38(7):848-856.

**[Response Ends]**

**1b.01. Briefly explain the rationale for this measure.**

*Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.*

**[Response Begins]**

The goal of this measure is to improve the process for diagnosis and treatment of urinary tract infection (UTI). Literature has demonstrated that while UTI is one of the most common infectious etiologies for which patients are hospitalized, it is often inappropriately diagnosed, resulting in inappropriate antibiotic administration and delay in diagnosis of a true underlying condition. The implications of inappropriate antibiotics are well described and include risks of antibiotic-associated adverse events such as *Clostridioides difficile* infection, prolonged length of hospital stay, and antimicrobial resistance, all of which can increase patient morbidity and mortality. Missed or delayed diagnosis of a true underlying condition is equally troubling, as data suggest that diagnostic error results in the highest morbidity, mortality, and malpractice cost of any medical error. Through adoption of this measure, we anticipate a decrease in inappropriate diagnosis of UTI, a decrease in unnecessary antibiotic use, and improved patient outcomes.

**[Response Ends]**

**1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.**

*Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

**[Response Begins]**

Data below are from 7/1/2017-3/31/2020 across 49 acute care hospitals in the state of Michigan. This includes 13,805 patients treated for UTI, of whom 23.2% (3,197) were inappropriately diagnosed with UTI. For data of scores over time, we do not report 2020 data as it only includes a single quarter (ending 3/31/2020).

Here, we divided all 49 hospitals each year into performance deciles with decile 1 representing the top performing hospitals. Scores or the percentage of patients treated for pneumonia who were considered inappropriately diagnosed with UTI are then reported by decile, first giving mean (standard deviation [SD]) then providing median (inter-quartile range [IQR]) data.

**Table 1**. Mean (SD) percent of cases inappropriately diagnosed with UTI (i.e., "score") by Year; N=49 hospitals

| Decile | 2017; mean (SD) | 2018; mean (SD) | 2019; mean (SD) |
|---|---|---|---|
| **1 (best performing)** | **10.6 (3.2)** | **11.5 (2.3)** | **5.3 (4.1)** |
| **2** | **19.0 (1.4)** | **15 (0.6)** | **10.7 (0.5)** |
| **3** | **22.3 (0.2)** | **18.2 (0.8)** | **13.3 (1.1)** |
| **4** | **23.5 (0.4)** | **21.4 (0.6)** | **16.8 (0.7)** |
| **5** | **24.9 (0.5)** | **23.6 (0.9)** | **18.8 (0.3)** |
| **6** | **27.6 (0.9)** | **26.8 (0.6)** | **20.0 (0.5)** |
| **7** | **30.5 (0.8)** | **28.7 (0.9)** | **23.4 (1.2)** |
| **8** | **34.0 (1.8)** | **32.1 (0.5)** | **26.5 (0.7)** |
| **9** | **40.2 (2.5)** | **35.5 (0.4)** | **28.5 (1.2)** |
| **10 (worst performing)** | **60.7 (22.9)** | **41.7 (7.2)** | **32.4 (2.2)** |

Mean (SD) percent of cases inappropriately diagnosed with UTI trended downward from 2017 to 2019 in all deciles.

*2020 includes only 1 quarter of data and thus is not reported in the time trend above.

**Table 2**. Median (IQR) percent of cases inappropriately diagnosed with UTI (i.e., "score") by Year; N=49 hospitals

| Decile | 2017; median (IQR) | 2018; median (IQR) | 2019; median (IQR) |
|---|---|---|---|
| 1 (best performing) | 11.2 (8.1, 1.3) | 11.5 (9.8, 13.1) | 6.0 (2.1, 8.5) |
| 2 | 18.4 (18.3, 19.7) | 14.9 (14.6, 15.5) | 10.8 (10.3, 11.0) |
| 3 | 22.3 (22.1, 22.5) | 18.3 (17.6, 18.9) | 13.4 (12.5, 14.1) |
| 4 | 23.4 (23.3, 23.7) | 21.7 (20.9, 21.9) | 16.8 (16.3, 17.4) |
| 5 | 25.0 (24.6, 25.2) | 23.5 (22.9, 24.4) | 18.8 (18.6, 19.0) |
| 6 | 27.6 (27.0, 28.1) | 27.0 (26.4, 27,1) | 20.0 (19.6, 20.5) |
| 7 | 30.0 (30.0, 30.7) | 28.4 (28.3, 28.6) | 23.2 (22.6, 24.2) |
| 8 | 33.3 (33.3, 33.9) | 32.0 (31.7, 32.5) | 26.3 (26.0, 26.9) |
| 9 | 40.4 (38.3, 42.2) | 35.5 (35.2, 35.8) | 28.2 (27.8, 29.2) |
| 10 (worst performing) | 53.8 (46.7, 60.0) | 40.0 (38.0, 40.3) | 31.7 (30.9, 33.8) |

Median (IQR) percent of cases inappropriately diagnosed with UTI trended downward from 2017 to 2019 in all deciles.

*2020 includes only 1 quarter of data and thus is not reported in the time trend above.

**[Response Ends]**

**1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.**

**[Response Begins]**

N/A

**[Response Ends]**

**1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.**

*Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

**[Response Begins]**

Data below are from 7/1/2017-3/31/2020 across 49 acute care hospitals in the state of Michigan. This includes 13,805 patients treated for UTI, of whom 23.2% (3,197) were inappropriately diagnosed with UTI.

Here, we report the demographics for patients with UTI as compared to the demographics of patients inappropriately diagnosed with UTI. We also compare demographics of those inappropriately diagnosed in 2017 to those inappropriately diagnosed in 2020. All comparisons were conducted using chi-squared tests.

**Table 3**. Demographics of UTI cohort and inappropriately diagnosed patients, Year 2017

| Variable | UTI, N=1929; % (N) | Inappropriate Diagnosis of UTI, N=752; % (N) | P-value |
|---|---|---|---|
| Medicaid | 9.9% (191) | 6.0% (45) | <.001 |
| Medicare | 75.4% (1455) | 83.1% (625) | * |
| Private Insurance | 14.7% (284) | 10.9% (82) | * |
| Female | 69.3% (1402) | 79.2% (614) | <.001 |
| Male | 30.7% (622) | 20.8% (161) | * |
| Race Black | 19.7% (399) | 18.2% (141) | 0.505 |
| Race Other[b] | 3.9% (78) | 3.4% (26) | * |
| Race White | 76.4% (1548) | 78.5% (609) | * |
| Age 65 years or older | 71.3% (1443) | 80.7% (626) | <.001 |
| Age < 65 years | 28.7% (582) | 19.3% (150) | * |

Demographic comparisons of the UTI cohort to those inappropriately diagnosed with UTI in 2017 indicate significant differences by payer, gender, and age. Patients inappropriately diagnosed with UTI were more likely to have medicare insurance (vs. private or Medicaid) compared to patients with UTI. Compared to patients with UTI, patients inappropriately diagnosed with UTi were more likely to be women and more likely to be older than 65 years. There were no differences by race.

*cell intentionally left empty

[a] P-value compares demographics of patients with UTI to those inappropriately diagnosed with UTI using chi-squared tests. P<0.05 considered significant.

[a]"other" race includes American Indian or Alaskan Native, Arab and Chaldean Ancestries, Asian, Native Hawaiian or Pacific Islander, Other (i.e., if patient demographic information indicates the patient is a race other than what is listed above), and Unknown (i.e., if patient's race is not indicated in the medical record).

**Table 4**. Demographics of entire UTI cohort and inappropriately diagnosed patients, Q1 2020

| Variable | UTI, N=561; % (N) | Inappropriate Diagnosis of UTI, N=140; % (N) | P-value |
|---|---|---|---|
| Medicaid | 10.1% (60) | 8.6% (12) | 0.020 |
| Medicare | 78.9% (470) | 87.9% (123) | * |
| Private Insurance | 11.1% (66) | 3.6 % (5) | * |
| Female | 69.3% (446) | 75.8% (113) | 0.112 |
| Male | 30.8% (198) | 24.2% (36) | * |
| Race Black | 22.3% (144) | 22.2% (33) | 0.933 |
| Race Other[b] | 5.4% (35) | 4.7% (7) | * |
| Race White | 72.3% (466) | 73.2% (109) | * |
| Age 65 years or older | 73.8% (476) | 77.9% (116) | 0.306 |
| Age < 65 years | 26.2% (169) | 22.2% (33) | * |

Demographic comparisons of the UTI cohort to those inappropriately diagnosed with UTI in quarter 1 of 2020 indicate significant differences by payer. Patients inappropriately diagnosed with UTI were more likely to have Medicare insurance (vs. private or Medicaid) compared to patients with UTI. There were no differences between patients with UTI and those inappropriately diagnosed with UTI by gender, race, or age.

*cell intentionally left empty

Abbreviations: Q1: quarter 1

[a] P-value compares demographics of patients with UTI to those inappropriately diagnosed with UTI using chi-squared tests. P<0.05 considered significant.

[a]"other" race includes American Indian or Alaskan Native, Arab and Chaldean Ancestries, Asian, Native Hawaiian or Pacific Islander, Other (i.e., if patient demographic information indicates the patient is a race other than what is listed above), and Unknown (i.e., if patient's race is not indicated in the medical record).

**Table 5**. Trends in demographics of patients inappropriately diagnosed with UTI; 2017 vs. Q1 2020

| Variable | 2017 Inappropriately Diagnosed with UTI, N=752; % (N) | Q1, 2020 Inappropriately Diagnosed with UTI, N=140; % (N) | P-value |
|---|---|---|---|
| Medicaid | 6.0% (45) | 8.6% (12) | 0.020 |
| Medicare | 83.1% (625) | 87.9% (123) | * |
| Private Insurance | 10.9% (82) | 3.6 % (5) | * |
| Female | 79.2% (614) | 75.8% (113) | 0.355 |
| Male | 20.8% (161) | 24.2% (36) | * |
| Race Black | 18.2% (141) | 22.2% (33) | 0.342 |
| Race Other[b] | 3.4% (26) | 4.7% (7) | * |
| Race White | 78.5% (609) | 73.2% (109) | * |

| Variable | 2017 Inappropriately Diagnosed with UTI, N=752; % (N) | Q1, 2020 Inappropriately Diagnosed with UTI, N=140; % (N) | P-value |
|---|---|---|---|
| Age 65 years or older | 80.7% (626) | 77.9% (116) | 0.429 |
| Age < 65 years | 19.3% (150) | 22.2% (33) | * |

A higher percentage of inappropriately diagnosed cases were seen for patients with Medicare and Medicaid compared to private insurance (P=0.02). Other demographics (gender, race, and age) of patients inappropriately diagnosed with UTI were not different between all of 2017 to quarter 1 of 2020 (P=0.34-0.43).

*cell intentionally left empty

Abbreviations: Q1: quarter 1

[a]P-value compares demographics of patients inappropriately diagnosed with UTI in 2017 to those inappropriately diagnosed with UTI in Q1 of 2020 using chi-squared tests. P<0.05 considered significant.

[a]"other" race includes American Indian or Alaskan Native, Arab and Chaldean Ancestries, Asian, Native Hawaiian or Pacific Islander, Other (i.e., if patient demographic information indicates the patient is a race other than what is listed above), and Unknown (i.e., if patient's race is not indicated in the medical record).

**[Response Ends]**

**1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.**

**[Response Begins]**

N/A

**[Response Ends]**

## Criteria 2: Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

---

**sp.01. Provide the measure title.**

*Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like ).*

**[Response Begins]**

Inappropriate diagnosis of urinary tract infection (UTI) in hospitalized medical patients; Abbreviated form: Inappropriate diagnosis of UTI

**[Response Ends]**

**sp.02. Provide a brief description of the measure.**

*Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).*

**[Response Begins]**

The inappropriate diagnosis of UTI in hospitalized medical patients (or "Inappropriate Diagnosis of UTI") measure is a process measure that evaluates the annual proportion of hospitalized adult medical patients treated for UTI who do not meet diagnostic criteria for UTI (thus are inappropriately diagnosed and overtreated).

**[Response Ends]**

**sp.04. Check all the clinical condition/topic areas that apply to your measure, below.**

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Surgery: General*

**[Response Begins]**

Genitourinary (GU): Urinary Tract Injection (UTI)

Infectious Diseases (ID)

**[Response Ends]**

**sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.**

**[Response Begins]**

Safety

Safety: Healthcare Associated Infections

Safety: Overuse

**[Response Ends]**

**sp.06. Select one or more target population categories.**

*Select only those target populations which can be stratified in the reporting of the measure's result.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Populations at Risk: Populations at Risk*

**[Response Begins]**

Adults (Age >= 18)

Elderly (Age >= 65)

**[Response Ends]**

**sp.07. Select the levels of analysis that apply to your measure.**

*Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

**[Response Begins]**

Facility

**[Response Ends]**

**sp.08. Indicate the care settings that apply to your measure.**

*Check ONLY the settings for which the measure is SPECIFIED and TESTED.*

**[Response Begins]**

Inpatient/Hospital

**[Response Ends]**

**sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.**

*Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".*

**[Response Begins]**

https://mi-hms.org/inappropriate-diagnosis-urinary-tract-infection-uti-hospitalized-medical-patients

**[Response Ends]**

**sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.**

*Attach an excel or csv file; if this poses an issue, contact staff . Provide descriptors for any codes. Use one file with multiple worksheets, if needed.*

**[Response Begins]**

Available in attached Excel or csv file

**[Response Ends]**

Attachment: 3690_Data_Dictionary_UTI_Measure_3.22.22.xlsx

**sp.12. State the numerator.**

*Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).*

*DO NOT include the rationale for the measure.*

**[Response Begins]**

The measure quantifies adult, hospitalized medical patients inappropriately diagnosed with UTI. Here, inappropriate diagnosis is defined as patients treated with antibiotics for UTI who do not meet diagnostic criteria for UTI. Patients were considered inappropriately diagnosed if they received antibiotic therapy for a UTI but did not have at least one sign or symptom of a UTI.

**[Response Ends]**

**sp.13. Provide details needed to calculate the numerator.**

*All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

Patients in the numerator include those that received antibiotics for a UTI but did not have ≥1 sign or symptom of a UTI.

- Minor numerator exclusions:
    - Those with a blood culture positive for a pathogenic bacteria (1.8% [91/4961])

Signs (e.g., fever) and symptoms (e.g., dysuria) of UTI are found in the attached excel file. Abstractors are asked to review the medical record for documentation of any signs or symptoms the day prior to obtaining a urine culture (referred to as day -1), the day of the urine culture (day 0), or the two days following the urine culture (days 1, 2). Any combination of 1 or more symptoms at any point in this time frame is required to be considered appropriately diagnosed. The exception is patients with new onset mental status changes. Consistent with recent IDSA guidelines, patients with new onset mental status changes must also have signs of a systemic infection (i.e., leukocytosis, hypotension, or ≥ 2 systemic inflammatory

response syndrome [SIRS] criteria) to be considered a UTI. Any patients without signs and symptoms of a UTI are considered inappropriately diagnosed and placed in the numerator.

**[Response Ends]**

**sp.14. State the denominator.**

*Brief, narrative description of the target population being measured.*

**[Response Begins]**

The denominator includes all adult, general care, immunocompetent, medical patients hospitalized and treated for UTI who do not have a concomitant infection.

**[Response Ends]**

**sp.15. Provide details needed to calculate the denominator.**

*All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

The denominator includes all sampled patients eligible for abstraction during the measure period (typically annual measurement). To be considered "treated for a UTI," a patient had to: a) have a positive urine culture, b) receive antibiotic therapy, and c) not have a concomitant infection. Please see excel file (inclusion criteria tab) for detailed operationalized definitions.

***Inclusion criteria:***

- Adult patient admitted and discharged from the participating hospital
- With a positive urine culture (except for excluded organisms listed in data dictionary) during hospitalization.
- Admitted to a general care medicine service
- Received any eligible antibiotic during the symptom collection window (day -1, 0, 1, 2, where day 0 = day of first positive urine culture)
- Immunocompetent (allowing for mild immune suppression)
- Do not have a concomitant infection (e.g., COVID-19, antibiotic treatment for unrelated infection or prophylaxis)
- Have normal urinary anatomy

**[Response Ends]**

**sp.16. Describe the denominator exclusions.**

*Brief narrative description of exclusions from the target population.*

**[Response Begins]**

***Exclusion Criteria:***

- Left against medical advice or refused medical care
- Admitted on hospice
- Pregnant or breastfeeding
- Spinal cord injury
- UTI-related complication (e.g., perinephric abscess)
    - Operationalized as >14 days of antibiotics at discharge

**[Response Ends]**

**sp.17. Provide details needed to calculate the denominator exclusions.**

*All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

Inclusion and exclusion codes and criteria are provided in the attached excel file.

**[Response Ends]**

**sp.18. Provide all information required to stratify the measure results, if necessary.**

*Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.*

**[Response Begins]**

N/A

**[Response Ends]**

**sp.19. Select the risk adjustment type.**

*Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.*

**[Response Begins]**

No risk adjustment or risk stratification

**[Response Ends]**

**sp.20. Select the most relevant type of score.**

*Attachment: If available, please provide a sample report.*

**[Response Begins]**

Rate/proportion

**[Response Ends]**

**sp.21. Select the appropriate interpretation of the measure score.**

*Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score*

**[Response Begins]**

Better quality = Lower score

**[Response Ends]**

**sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.**

*Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.*

**[Response Begins]**

The measure estimates hospital-level inappropriate diagnosis of UTI. If the hospital has elected to sample patients, they will generate a sample by first identifying all hospitalized patients with a positive urine culture (using institutional definition of positive) during that month or quarter (based on whether they elect to sample monthly or quarterly). Next, they will apply electronic inclusion criteria (medicine admission, antibiotic receipt during window period [day -1 to day +2]) to either their quarterly or monthly patient sample. The resulting list will be randomized, and patients screened in order of randomization. First, patients are screened for inclusion in the denominator. All adult, general care, medical patients hospitalized and treated for UTI are potentially eligible. If the patient meets eligibility criteria and does not have any exclusions, they are placed in the denominator. Patients automatically excluded from the numerator are those with blood cultures positive for a pathogenic organism. Patients are then assessed for whether they meet diagnostic criteria for UTI (i.e., do they have at least one sign or symptom of a UTI). If a patient does NOT meet diagnostic criteria they are placed in the numerator. A lower score is considered better diagnostic quality for UTI.

**[Response Ends]**

**sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.**

**[Response Begins]**

**Sampling**: Hospitals have the option to sample from their population or submit their entire population. Hospitals also have the option to sample quarterly or monthly. Over the entire year, 59 cases are recommended for the denominator. Thus, hospitals whose Initial Patient Population size is less than or equal to the minimum number of cases per quarter (N=15) or month (N~5) for the measure should not sample and rather, should include all cases. A hospital may choose to use a larger sample size than is required.

**Sampling Procedures:**

Potentially eligible patient lists should be reviewed monthly or quarterly (as desired). Lists will be determined by the ability of the facility; however, we suggest electronically including the following criteria:

- Initial sample based on positive urine culture
- Exclude patients who did not receive antibiotics during hospitalization (if able, can refine to day -1 to day +2 with day 0 being date of urine culture collection)
- Exclude patients admitted to a non-medicine service
- Exclude patients admitted to intensive care

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting to sample cases MUST submit AT LEAST the minimum required sample size.

Eligible lists should then be randomized and reviewed in order until the desired number of cases is included (~5/month or ~15 per quarter).

**Minimum Sample Size:**

Using the Spearman Brown prophecy, we evaluated the number of cases needed to reach each reliability threshold:

Table 1. Number of annual cases needed to achieve each reliability threshold.

| **Reliability** | **Number of annual cases needed** |
|---|---|
| 0.6 | 22 |
| 0.7 | 35 |
| 0.8 (standard) | **59** |
| 0.9 | 132 |

Based on these data, for a desired reliability of 0.8, each hospital would need to abstract 59 cases annually or ~5 cases per month.

**[Response Ends]**

**sp.28. Select only the data sources for which the measure is specified.**

**[Response Begins]**

Electronic Health Data

Electronic Health Records

Other (specify)

Chart Review

**[Response Ends]**

**sp.29. Identify the specific data source or data collection instrument.**

*For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.*

**[Response Begins]**

Electronic medical record data. The data collection instrument is provided. Those interested in using our online REDCap tool may also contact us directly to coordinate.

**[Response Ends]**

**sp.30. Provide the data collection instrument.**

**[Response Begins]**

Available in attached appendix in Question 1 of the Additional Section

**[Response Ends]**

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

• Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.

• All required sections must be completed.

• For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.

• If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.

• An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.

• Contact NQF staff with any questions. Check for resources at the

Submitting Standards webpage .

• For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the

2021 Measure Evaluation Criteria and Guidance .

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

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) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

• an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

OR

• rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

***Definitions***

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:

***2021 Submission:***

Updated testing information here.

***2018 Submission:***

Testing from the previous submission here.

**2a.01. Select only the data sources for which the measure is tested.**

**[Response Begins]**

Electronic Health Data

Electronic Health Records

Other (specify)

Chart Review

**[Response Ends]**

**2a.02. If an existing dataset was used, identify the specific dataset.**

*The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).*

**[Response Begins]**

For reliability testing, we used data from the Michigan Hospital Medicine Safety Consortium (HMS). HMS is a collaborative quality initiative sponsored by Blue Cross Blue Shield of Michigan (https://mi-hms.org/ ). HMS includes 62 non-governmental hospitals throughout the state of Michigan. In July 2017, HMS hospitals joined in the "Antimicrobial Use Initiative" to collect patient-level data related to hospitalized, medical patients treated for urinary tract infection (UTI) (https://mi-hms.org/quality-initiatives/antimicrobial-use-initiative ).[1,2]

For all analyses included in this measure submission, data from HMS are censored as of March 31, 2020, at which time 49 hospitals had contributed data to the dataset.

The dataset includes chart abstracted data, such as:

- Patient demographics (e.g., age, admission, and discharge dates)
- Positive urine culture information (e.g., organisms)
- Presence of signs or symptoms of a UTI within the period of the day prior to the urine culture being collected through two days after urine culture being collected (day -1 to +2 where the urine culture collection date is day 0)
    - Physical exam findings (e.g., costovertebral angle tenderness)
    - Vital signs (e.g., fever)
    - Documented symptoms (e.g., dysuria)
    - Laboratory findings (e.g., leukocytosis)
- Antibiotic use during admission and on discharge
- Urinary catheter use
- Comorbidities – including diabetes, end stage renal disease (ESRD), dementia, admission from a skilled nursing facility/long term care facility
- 30-day adverse events (emergency department visit, mortality, *Clostridioides difficile* infection, antibiotic associated side effects) documented in the medical record
- 30-day adverse events collected via telephone interview (conducted 30-days post discharge)

References:

[1] Petty LA, Vaughn VM, Flanders SA, et al. Risk Factors and Outcomes Associated With Treatment of Asymptomatic Bacteriuria in Hospitalized Patients. *JAMA Intern Med.* 2019;179(11):1519–1527.

[2] Petty LA, Vaughn VM, Flanders SA, et al. Assessment of Testing and Treatment of Asymptomatic Bacteriuria Initiated in the Emergency Department. Open Forum Infect Dis. 2020 Nov 3;7(12):ofaa537.

**[Response Ends]**

**2a.03. Provide the dates of the data used in testing.**

*Use the following format: "MM-DD-YYYY - MM-DD-YYYY"*

**[Response Begins]**

07-01-2017 - 03-31-2020

**[Response Ends]**

**2a.04. Select the levels of analysis for which the measure is tested.**

*Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

**[Response Begins]**

Facility

**[Response Ends]**

**2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).**

*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.*

**[Response Begins]**

**Table 1. Characteristics of Participating Hospitals**

| Hospital Characteristic | HMS Hospitals n=49; n (%) | All Michigan Hospitals[1] n=127; n (%) |
|---|---|---|
| Academic Hospital[1] | 40 (82%) | 74 (58%) |
| Location[2,3] | * | * |
| Metropolitan | 40 (82%) | 71 (56%) |
| Micropolitan | 8 (16%) | 24 (19%) |

| Hospital Characteristic | HMS Hospitals n=49; n (%) | All Michigan Hospitals[1] n=127; n (%) |
|---|---|---|
| Rural | 1 (2%) | 32 (25%) |
| Profit Type[2] | * | * |
| Non-Profit | 45 (92%) | 116 (59%) |
| For profit | 4 (8%) | 9 (33%) |
| Government | 0 (0%) | 2 (2%) |
| Bed Size (Staffed beds)[4] | * | * |
| ≤50 | 2 (4%) | 46 (36%) |
| 51-100 | 4 (8%) | 21 (17%) |
| 101-200 | 9 (18%) | 16 (13%) |
| >200 | 34 (69%) | 44 (35%) |

*Cells intentionally left empty

Data compiled from the following sources:

[1] List of Michigan Hospitals compiled from the Michigan Health & Hospital Association[§]

mha.org/about/our-hospitals Accessed January 3, 2022

[2] U.S. Census Bureau, Michigan: 2020 Core Based Statistical Areas and Counties

https://www2.census.gov/programs-surveys/metro-micro/reference-maps/2020/state-maps/26_Michigan_2020.pdf

[3] U.S. Census Bureau, Core based statistical areas (CBSAs), metropolitan divisions, and combined statistical areas (CSAs)

https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html

[4] American Hospital Directory, Individual Hospital Statistics for Michigan

https://www.ahd.com/states/hospital_MI.html

[§]The following types of hospitals were excluded:

- Children's hospitals
- Long-term acute care hospitals
- Psychiatric/mental health/substance abuse hospitals
- Rehabilitation hospitals
- Surgical hospitals
- Those providing only specialty services (i.e., cardiac hospital)

**[Response Ends]**

**2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.**

*If there is a minimum case count used for testing, that minimum must be reflected in the specifications.*

**[Response Begins]**

Between 7/1/2017 and 3/31/2020 there were 13,805 hospitalized patients treated for UTI across 49 HMS hospitals. All 13,805 patients were used to test validity and reliability of the inappropriate diagnosis of UTI measure. Of the 13,805

patients treated for UTI, 23.2% (3,197) were assessed to be inappropriately diagnosed with UTI. Reliability and validity were assessed at the hospital level and validity was assessed at the encounter (i.e., patient) level. Descriptive characteristics of the entire UTI cohort are as follows:

Table 2. Descriptive characteristics of the entire UTI cohort, patients with appropriate diagnosis of UTI, and patients with inappropriate diagnosis of UTI

| Characteristic | Entire UTI Cohort, n (%) | Appropriate Diagnosis, n (%) | Inappropriate Diagnosis, n (%) |
|---|---|---|---|
| *Gender* | * | * | * |
| Male | 4097 (29.7%) | 3311 (31.2%) | 786 (24.6%) |
| Female | 9702 (70.3%) | 7292 (68.7%) | 2410 (75.4%) |
| *Race* | * | * | * |
| White | 10257 (74.3%) | 7885 (74.3%) | 2372 (74.2%) |
| Black | 2945 (21.3%) | 2251 (21.2%) | 694 (21.7%) |
| Asian | 74 (0.5%) | 64 (0.6%) | 10 (0.3%) |
| American Indian | 37 (0.3%) | 26 (0.2%) | 11 (0.3%) |
| Native Islander | 22 (0.2%) | 18 (0.2%) | 4 (0.1%) |
| Other | 227 (1.6%) | 186 (1.8%) | 41 (1.3%) |
| Unknown | 190 (1.4%) | 143 (1.3%) | 47 (1.5%) |
| *Age (years)* | * | * | * |
| 18-30 | 494 (3.6%) | 445 (4.2%) | 49 (1.5%) |
| 31-40 | 453 (3.3%) | 399 (3.8%) | 54 (1.7%) |
| 41-50 | 624 (4.5%) | 515 (4.9%) | 109 (3.4%) |
| 51-60 | 1235 (8.9%) | 999 (9.4%) | 236 (7.4%) |
| 61-70 | 2435 (17.6%) | 1895 (17.9%) | 540 (16.9%) |
| 71-80 | 3463 (25.1%) | 2665 (25.1%) | 798 (25.0%) |
| 80-90 | 3709 (26.9%) | 2706 (25.5%) | 1003 (31.4%) |
| 91-100 | 1316 (9.5%) | 929 (8.8%) | 387 (12.1%) |
| 100+ | 76 (0.6%) | 55 (0.5%) | 21 (0.7%) |
| *Insurance Status* | * | * | * |
| Private | 1316 (9.5%) | 1077 (10.2%) | 239 (7.5%) |
| Medicare | 10165 (73.6%) | 7600 (71.6%) | 2565 (80.2%) |
| Medicaid | 1209 (8.8%) | 1012 (9.5%) | 197 (6.2%) |
| Uninsured | 114 (0.8%) | 105 (1.0%) | 9 (0.3%) |
| *Comorbidities* | * | * | * |
| Presence of urinary catheter | 1876 (13.6%) | 1426 (13.4%) | 450 (14.1%) |
| Renal disease | 5643 (40.9%) | 4303 (40.6%) | 1340 (41.9%) |

| Characteristic | Entire UTI Cohort, n (%) | Appropriate Diagnosis, n (%) | Inappropriate Diagnosis, n (%) |
|---|---|---|---|
| Liver disease | 811 (5.9%) | 636 (6.0%) | 175 (5.5%) |
| Congestive heart failure | 3241 (23.5%) | 2403 (22.7%) | 838 (26.2%) |
| Chronic obstructive pulmonary disease | 2507 (18.2%) | 1889 (17.8%) | 618 (19.3%) |
| Home oxygen | 619 (4.5%) | 457 (4.3%) | 162 (5.1%) |
| Structural lung disease | 0 (0%) | 0 (0%) | 0 (0%) |
| Current/Former smoker | 6489 (47%) | 5111 (48.2%) | 1378 (43.1%) |
| Cancer | 2778 (20.1%) | 2143 (20.2%) | 635 (19.9%) |
| Immune compromise | 95 (0.7%) | 74 (0.7%) | 21 (0.7%) |
| Diabetes mellitus | 5331 (38.6%) | 4111 (38.8%) | 1220 (38.2%) |
| Sepsis | 3774 (27.3%) | 3551 (33.5%) | 223 (7%) |
| Severe Sepsis | 339 (2.5%) | 339 (3.2%) | 0 (0%) |

*Cells intentionally left empty

Hospitals within HMS use the following case identification strategy to determine patients to abstract for HMS:

- Data collection involves abstraction of eligible cases every two weeks.
- To minimize sampling bias, abstractors are expected to select cases from every day during a two-week period, including weekends.
- The list of cases eligible for abstraction is created using the below protocol
    - For each two-week period, a list of patients admitted to all medical services is created
        - For inappropriate diagnosis of UTI, this list is generally a list of all positive urine cultures
        - If possible, hospitals apply additional electronic filters to the dataset to screen for inclusion/exclusion criteria. For example, they may exclude patients from the "inappropriate diagnosis of UTI" list if they also had a discharge diagnosis of pneumonia or were cared for on a non-medicine service.
        - All inclusion/exclusion criteria that are not electronically applied prior to list generation will require manual screening during case review
        - The list of potentially eligible patients is then organized chronologically by date and time of discharge.
        - For each discharge day, the first patient on the chronological list is reviewed for inclusion. If excluded, the next patient is reviewed.
        - This process is repeated, with patients reviewed from the chronological list ensuring that cases are distributed evenly across the two-week timeframe – meaning there are discharge dates across all days of the week – until all cases are identified and abstracted.

We do not report encounter-level reliability as we report encounter-level validity. Please see the validity documents for additional information.

**[Response Ends]**

**2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.**

**[Response Begins]**

All data analysis was performed on the same dataset.

Table 3. Description of samples utilized to determine hospital-level and encounter-level reliability and empirical validity

| Type of Testing | Sample Utilized |
|---|---|
| Hospital-Level Reliability and Empirical Validity[1] | Entire HMS UTI Dataset (based on case identification protocol outlined in 2a.06) |
| Encounter-Level Reliability[1] | *Assessment of Effect of Abstraction Errors*: Review of a random, consecutive subset of 50 encounters within the cohort, representing cases from 29 of 46 participating hospitals.<br>*Structured Implicit Case Review*: Seventeen cases, pseudo-randomly selected, for in-depth review by 2-4 physicians to confirm case classification (appropriate versus inappropriate diagnosis) |

[1]Please see validity documents for further information.

**[Response Ends]**

**2a.08. List the social risk factors that were available and analyzed.**

*For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.*

**[Response Begins]**

As this is a process measure, no risk adjustment was performed (including for social factors).

**[Response Ends]**

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

**2a.09. Select the level of reliability testing conducted.**

*Choose one or both levels.*

**[Response Begins]**

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

Accountable Entity Level (e.g., signal-to-noise analysis)

**[Response Ends]**

**2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.**

*Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.*

**[Response Begins]**

**Patient or Encounter Level**

Please see validity testing section for encounter-level validity.

**Accountable Entity Level**

Signal-to-noise analysis was performed using a mixed-effect logistic model ran as an empty model such that the only effects in the model were the overall intercept and the hospital specific intercepts. This model enabled for the calculation of the hospital variance (signal), the total variance, and the within hospital variance (noise). Based on the hospital variance and the within hospital variance, an intraclass correlation was calculated. The intraclass correlation was utilized within the Spearman Brown formula in two ways: (A) to calculate the reliability for the entire hospital cohort using the median number of case abstractions for the cohort and (B) to understand minimum case abstracts necessary to achieve predetermined reliability thresholds of 0.6, 0.7, 0.8, and 0.9.

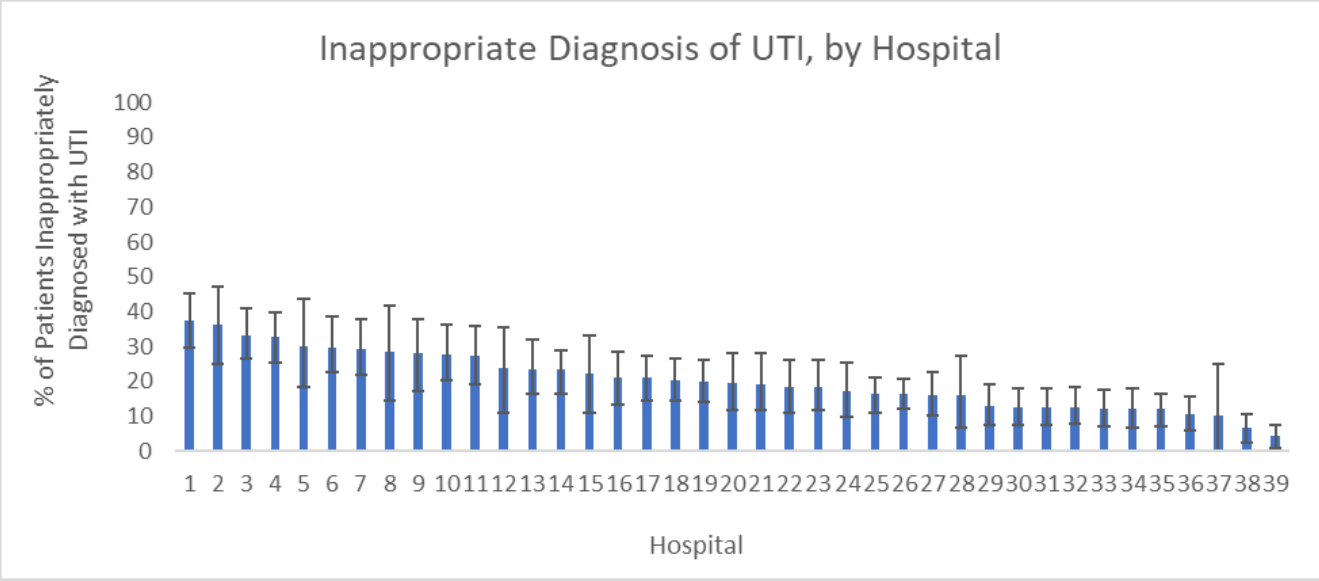
**[Response Ends]**

**2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?**

*For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria ).*

**[Response Begins]**

Figure 1. Distribution of percentage of patients inappropriately diagnosed with UTI by hospital with 95% confidence intervals is demonstrated below. These data are based on the 4 quarters preceding March 2020 and include only hospitals that provided data during all four quarters.




From these data, we were able to calculate the following:

Hospital Variance (signal): 0.225271

Total Variance: 3.5151414

Within Hospital Variance (noise): 3.28987

Based on this information, an intraclass correlation (ICC) was calculated. This ICC represents the reliability of the cohort if a single measurement (case abstraction) per hospital were included.

ICC=0.225271/(0.225271+3.28987)=0.225271/3.5151414=0.0641

A. The Spearman Brown Prophecy allows to an estimation of reliability after adjusting the number of measurements. We can use this formula to estimate the reliability of the measure within the cohort after adjusting the input (in this case the number of case abstractions per site).[1,2] The Spearman Brown Formula states the following:

$Reliability_{new}$ = (n*r)/(1+[n-1]*r) where n is the number of inputs and r is the prior reliability.

Adapting to the formula to our variables suggests the following:

$Reliability_{new}$ = (number of case reviews*ICC)/(1+[number of case reviews-1]* ICC)

The median case abstraction counts for the entire cohort was applied to the Spearman Brown Formula to obtain the overall reliability for the cohort.

Median case abstractions: 133 (IQR 92-154)

Reliability: (133*0.0640859)/(1+(133-1)*0.0640859)=0.901

1. Spearman, C. (1910), Correlation Calculated From Faulty Data. *British Journal of Psychology*, 1904-1920, 3: 271-295.

2. Warrens MJ. Transforming intraclass correlation coefficients with the Spearman-Brown formula. *J Clin Epidemiol*. 2017 May;85:14-16

B. The ICC was then applied to the Spearman Brown Formula to calculate the minimum number of cases to achieve pre-specified reliability thresholds based on the outcome distribution of the entire cohort.

Table 1. Number of annual cases needed to achieve each reliability threshold.

| **Reliability** | **Number of annual cases needed** |
|---|---|
| 0.6 | 22 |
| 0.7 | 35 |
| 0.8 (standard) | **59** |
| 0.9 | 132 |

To achieve a desired reliability of 0.8, each hospital would need to abstract 59 cases annually.

**[Response Ends]**

**2a.12. Interpret the results, in terms of how they demonstrate reliability.**

*(In other words, what do the results mean and what are the norms for the test conducted?)*

**[Response Begins]**

A. Based on signal-to-noise analysis, we found that reliability of the measure across the entire hospital cohort was strong (0.90), meeting the threshold for reliability for measures considered to be high stakes.

B. Using the current HMS cohort as a representative example, the minimum number of case abstracts per hospital per year to meet pre-specified reliability thresholds of 0.7 and 0.8 are highly attainable. Within a cohort of 40 HMS hospitals participating in 2019, 90% of hospitals were able to abstract the minimum of 59 cases to achieve 0.8 reliability. Of those

that could not abstract the required number of cases, hospital bed sizes were 49 beds, 68 beds, 75 beds, and 133 beds. Ninety-five percent of hospitals could abstract the 35 cases/year necessary to achieve 0.7 reliability, and all but one could reach the abstraction threshold for 0.6 reliability. Of the two hospitals unable to achieve abstraction thresholds for 0.7 reliability (75 beds and 133 beds), one hospital over-sampled casers for an alternative measure, and the other had challenges with data abstractor hiring. This cohort of 40 hospitals participating in 2019 was selected as this represented the last year of complete data collection prior to the COVID-19 pandemic.

**[Response Ends]**

**2b.01. Select the level of validity testing that was conducted.**

**[Response Begins]**

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

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)

**[Response Ends]**

**2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.**

*Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.*

**[Response Begins]**

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**) | 2019 IDSA Asymptomatic Bacteriuria Guidelines[1] | Initial basis for definitions |
| B. Face Validity-Expert Feedback | Data Design and Publications Committee and Michigan Hospital Medicine Safety Consortium (HMS) 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 |

| Process | Description (stage of measure) | Results | Interpretation |
|---|---|---|---|
| **During Measure Development** | * | * | * |
| **During Early Years (2017-2019) of Measure Use** | * | * | * |
| C. Encounter-level Validity: Inappropriate Diagnosis Case Reporting | All inappropriately diagnosed cases reported to participating hospital **(Early Measure AND Current Measure as Specified)** | Minor adjustments based on feedback from real cases | 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 29 hospitals) to assess effect of any discrepancies on encounter-level validity **(Current Measure as Specified)** | Overall abstraction accuracy was 98.6%.<br>Two cases changed classification due to discrepancies noted in audit.<br>IRR: Kappa = 0.91<br>95% CI (0.78 – 1.00)<br>Strong to "almost perfect" reliability | Encounter-level validity is high with a "strong" to "almost perfect" reliability.<br>Data abstraction is typically accurate; what mistakes are made generally do not affect case classification. |
| E. Encounter-level Validity: Structured Implicit Case Review | 25 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 UTI] by HMS do you agree are [inappropriately diagnosed] (0-100%)?" **(Current Measure as Specified)** | Median: 90%<br>IQR: 80% to 97% | Most participating hospitals believed the measure was highly accurate |

| Process | Description (stage of measure) | Results | Interpretation |
|---|---|---|---|
| **During Measure Development** | * | * | * |
| G. Face Validity: National Expert Panel Feedback (N=11 experts) | Individuals representing 11 national organizations participated in 2-week online discussion of measure.<br>**(Current Measure as Specified)** | Survey Question:<br>"The inappropriate diagnosis of UTI measure as specified can be used to distinguish between better and worse quality hospitals." Likert (1=Strongly disagree, 5=Strongly agree)<br>9 respondents (82%) reported that they agreed/strongly agreed with this statement. | Measure with substantial face validity by TEP<br>Additional feedback to improve validity was 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 UTI measure was valid and important |
| I. Empirical Validity: Evaluated association with other measures of diagnostic quality | Evaluated association at hospital level between UTI inappropriate diagnosis and inappropriate diagnosis of community acquired pneumonia (CAP). **(Current Measure as Specified)** | Hospitals with higher rates of inappropriate diagnosis of UTI also had higher rates of inappropriate diagnosis of CAP; R=0.53<br>(i.e., moderate positive correlation) | Hospitals performing better on this measure were also better at appropriately diagnosing CAP |
| J. Empirical Validity: Evaluated association of inappropriate diagnosis of UTI with outcomes | Characterized antibiotic use in patients inappropriately diagnosed with UTI and the association of antibiotic use with adverse events after hospital discharge **(Current Measure as Specified)** | Median (IQR) 7 (4-9) unnecessary antibiotic days<br>Patients inappropriately diagnosed with UTI had an ~1 day longer length of stay after urine testing than those with asymptomatic bacteriuria (ASB) who were not treated with antibiotics (aRR: 1.37 [1.28-1.47]). | Inappropriate diagnosis of UTI is associated with unnecessary antibiotic use and longer hospitalizations |

*Cells intentionally left empty

**A. Face Validity Indicated by Established UTI Guidelines**

The initial definition of inappropriate diagnosis of UTI was derived from the "Clinical Practice Guideline for the Management of Asymptomatic Bacteriuria: 2019 Update by the Infectious Diseases Society of America."[1] Additional expert feedback and review helped refine measure development and design.

The 2019 Infectious Diseases Society of America Clinical Practice Guideline for the Management of Asymptomatic Bacteriuria (ASB) defines ASB as the following: "ASB is the presence of 1 or more species of bacteria growing in the urine at specified quantitative counts ($\geq 10^5$ colony-forming units [CFU]/mL or $\geq 10^8$ CFU/L), irrespective of the presence of pyuria, in the absence of signs or symptoms attributable to UTI."[1] This definition is consistent with our measure which defines inappropriate diagnosis of UTI as any patient treated for UTI that does not have signs or symptoms of a UTI. We

also use their criteria of when to treat altered mental status as a UTI: 1) when altered mental status occurs with other symptoms or 2) when patient has "other systemic signs of infection (e.g., fever or hemodynamic instability)."[1] 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.[2]

[1] Nicolle LE, Gupta K, Bradley SF, et al. Clinical Practice Guideline for the Management of Asymptomatic Bacteriuria: 2019 Update by the Infectious Diseases Society of America. *Clin Infect Dis.* 2019;68(10):e83-e110. doi:10.1093/cid/ciy1121.

[2] 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 these 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=3197 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 UTI (appropriate or inappropriate). These cases included 29 hospitals. Cases were scored based on 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 daily symptoms/signs, urinary catheter 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. Using standard methods, an inter-rater reliability was obtained to assess the difference between original case classification and true case classification after identifying data errors.

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

In 2020, we conducted structured implicit review of cases of inappropriate diagnosis of UTI by 2-4 physicians to confirm accurate case categorization. Cases were randomly selected from "gray areas" that had been brought up during initial measure development (e.g., patients with altered mental status). 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 UTI 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=40 hospitals)**

In October 2021 (after measure specifications had been finalized), we systematically assessed the perceived validity of the inappropriate diagnosis of UTI 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 UTI] by HMS do you agree are [inappropriately diagnosed] (0-100%)?"

**G. Face Validity: National Expert Panel Feedback (N=11 experts)**

Throughout measure development, we obtained expert and stakeholder input. In October 2021, we obtained formal expert feedback by holding a series of meetings over two-weeks with a national Technical Expert Panel (TEP). This TEP included representatives from societies and organizations who would potentially be impacted by the measure to provide feedback on the measure.

In alignment with the CMS Measures Management System guidance on TEPs,[3] 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, urologists, 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 |

| Organization/Institution | TEP Member |
|---|---|
| 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 |
| Emergency Medicine Physician, University of Wisconsin | Michael Pulia |
| American Urological Association (AUA) | Micheal Liss |

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

How much do you agree/disagree with the following statement?

1. "The inappropriate diagnosis of UTI 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.

2. Are there any key data elements you believe are missed or not accurately captured in the inappropriate diagnosis of UTI measure?

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

Finally, we solicited patient feedback through a Patient Engagement Panel in order to understand patient perspectives on the inappropriate diagnosis of UTI measure. 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 an infection 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 UTI 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 UTI with inappropriate diagnosis of CAP at the hospital level.

**J. Empirical Validity: Evaluated association of inappropriate diagnosis of UTI 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 UTI using descriptive statistics. Because duration was skewed, we report median (IQR/inter-quartile range) duration of antibiotic therapy.

Next, we compared outcomes in patients inappropriately diagnosed with UTI vs. those who had ASB but were not unnecessarily treated with antibiotics. Outcomes assessed included: 30-day mortality, 30-day hospital readmission, 30-day emergency department visit, discharge to post–acute care settings, *Clostridioides difficile* infection at 30 days, and

duration of hospitalization after urine testing. The association of inappropriate diagnosis with outcomes was assessed using logistic generalized estimating equation models, inverse probability of treatment weighted by baseline covariates identified to be significant in the bivariate and/or multivariate analysis, and other factors potentially associated with the outcome.

The results of this analysis were published in *JAMA Internal Medicine* in 2019 and are also shown below.[3]

[3] Petty LA, Vaughn VM, Flanders SA, et al. Risk Factors and Outcomes Associated With Treatment of Asymptomatic Bacteriuria in Hospitalized Patients. *JAMA Intern Med*. 2019. doi:10.1001/jamainternmed.2019.2871. PCMID: PMC6714039.

**[Response Ends]**

**2b.03. Provide the statistical results from validity testing.**

*Examples may include correlations or t-test results.*

**[Response Begins]**

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

In 2021, 50 cases were chronologically selected for detailed audit. Overall data element abstraction accuracy was 98.6%. When errors found through the data audit were corrected, there were two changes in case classification.

Table 3. Accuracy of abstractor vs auditor classification

| Abstractor Classification (original) | Auditor Classification (updated) | Number (n=50) |
|---|---|---|
| Inappropriate Diagnosis of UTI | Inappropriate Diagnosis of UTI | 14 |
| UTI | UTI | 34 |
| Inappropriate Diagnosis of UTI | UTI | 1 |
| UTI | Inappropriate Diagnosis of UTI | 1 |

Two cases changed classification due to discrepancies noted in audit. Thus, the IRR or Kappa was 0.91 (95% CI : 0.78 – 1.00) indicating strong to "almost perfect" reliability.

**E. Encounter-level Validity: Structured Implicit Case Review**

In 2020, 25 cases of inappropriate diagnosis of UTI underwent structured implicit case review by 2-4 physicians. **In 92% of cases (23/25) there was 100% agreement by reviewers that the cases represented 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 feedback during this case review process, we made minor refinements to our measure specifications, including refining our inclusion definitions. Specifically, two groups of patients would no longer be included: a) those who were never treated for a UTI even if symptomatic (because they are not inappropriately diagnosed), b) those who received antibiotics only outside of our symptom collection window (symptoms may have occurred later). We also added "hypogastric" as a synonym for "suprapubic" to ensure hypogastric pain was included as a UTI symptom.

**F. Face Validity: Feedback from HMS hospitals (N=40 hospitals)**

We systematically assessed the perceived validity (after finalization of measure specifications) of the inappropriate diagnosis of UTI measure by soliciting feedback from all participating HMS hospitals (N=40 hospitals) via the following question: "Approximately, what percentage of cases called ASB by HMS do you agree are inappropriately diagnosed with ASB (0-100%)." All hospitals (40/40) responded. Respondents were local leaders or quality champions for the measures.

Median: 90%  Inter-quartile range: 80% to 97%

**G. Face Validity: National Expert Panel Feedback**

Based on conversations held during our two-week online TEP, the 11 national experts who attended our TEP generally agreed with the face validity and operationalization of the overdiagnosis of UTI measure as currently specified. They believed that patients we identified as being inappropriately diagnosed were, in fact, inappropriately diagnosed. There were also some concerns about the use of the word "over-diagnosis" in the measure name. As a result, we changed the measure name to "inappropriate diagnosis" of UTI. There were no changes to measure specifications suggested by the TEP.

TEP Survey results:

**Table 4**. Distribution of TEP responses to ***Question #1***: "*The inappropriate diagnosis of UTI measure as specified can be used to distinguish between better and worse quality hospitals*."

| Rating | # of Responses (N=11) | Percent (%) | Cumulative Percent (%) |
|---|---|---|---|
| 5 (Strongly agree) | 1 | 9.1% | 9.1% |
| 4 (Agree) | 8 | 72.7% | 81.8% |
| 3 (Neutral) | 1 | 9.1% | 90.9% |
| 2 (Disagree) | 0 | 0.0% | 90.9% |
| 1 (Strongly disagree) | 1 | 9.1% | 100.0% |

**Table 5.** TEP responses to ***Question #2.*** "*What additional data would you like to see captured related to the inappropriate diagnosis of UTI? (free text)*" N=11 respondents (free text question)

| % of Responses N=11 | Response | Our Action/Response to Comment |
|---|---|---|
| 72.3% (8/11) | None or N/A | None. Confirmed validity of measurement. |
| 9.1% (1/11) | Duration of Antibiotic Treatment | Added data on duration of antibiotic treatment for patients inappropriately diagnosed with UTI to measure submission.<br>**Patients inappropriately diagnosed with UTI received a median (IQR) 7 (4-9) antibiotic days, all of which were unnecessary.** [3] |
| 9.1% (1/11) | Balancing Measure | Added additional resources on studies of underdiagnosis to measure submission |
| 9.1% (1/11) | Length of stay data | Added data on length of stay for patients inappropriately diagnosed with UTI to measure submission.<br>**Patients inappropriately diagnosed with UTI has a median (IQR) length of stay of 5 (4-7) days.**<br>Compared to patients with ASB not treated with antibiotics, **patients inappropriately diagnosed with UTI had a longer duration of hospitalization after urine testing** (4 vs. 3 days, adjusted relative risk 1.37).[3] |

**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."**

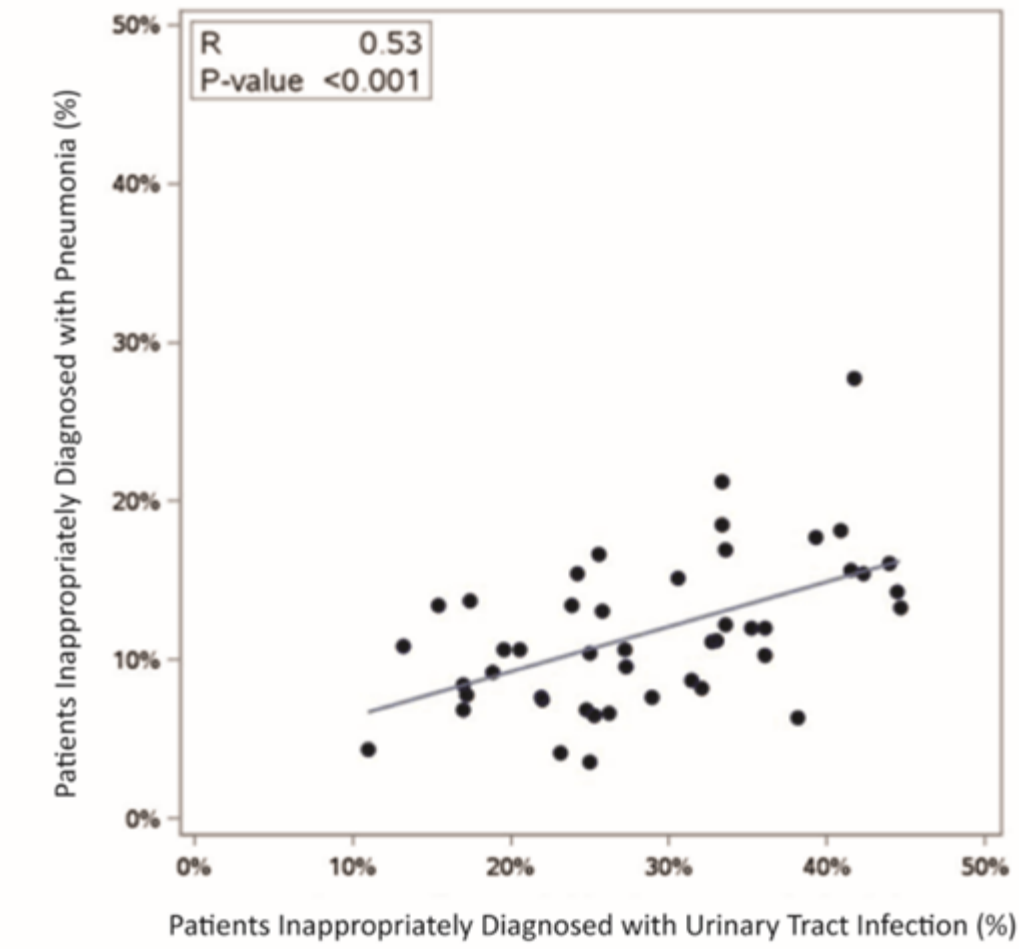
**"I would like it if they had a hospital rating… I think it would be beneficial, and I would really appreciate that. I feel that it would affect my decision of where I would go… it would definitely affect where I would guide my family or loved one to go."**

**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**

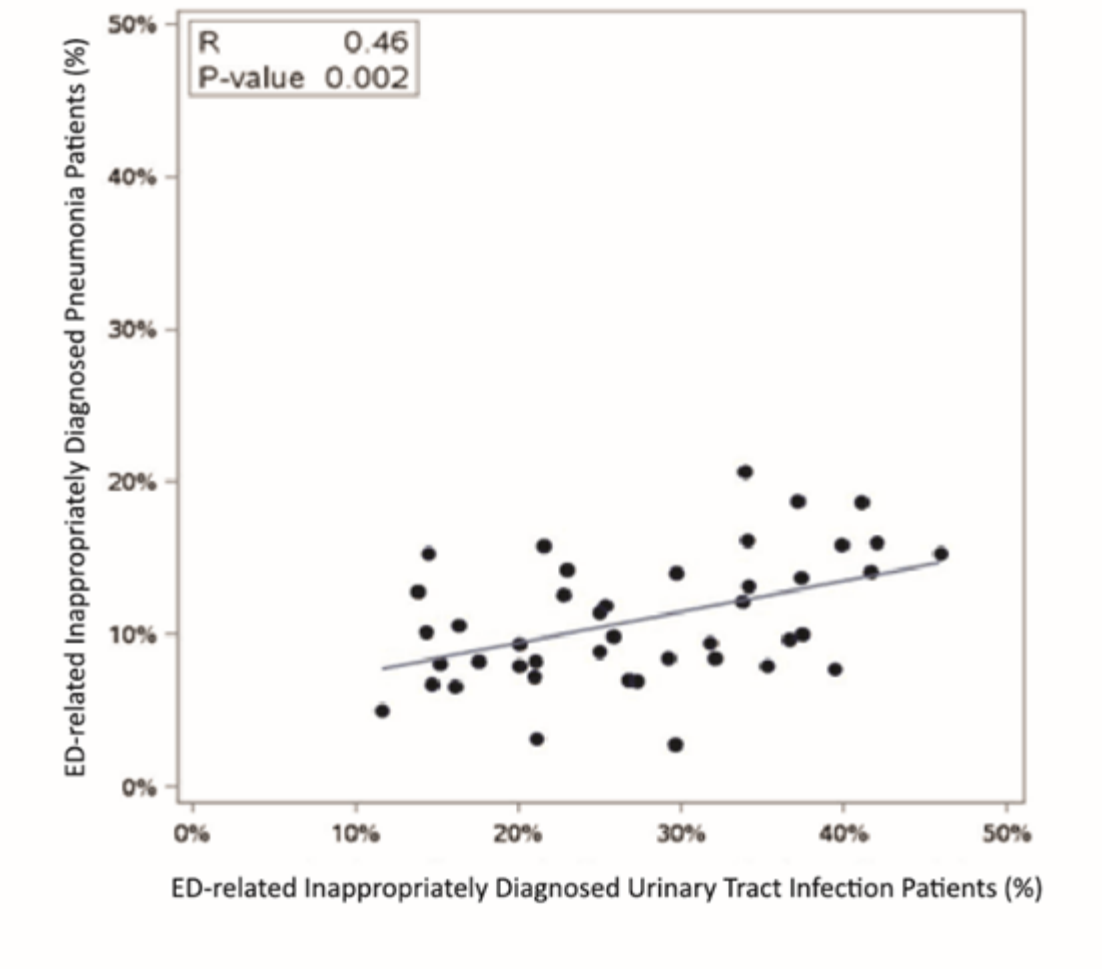
To address whether inappropriate diagnosis of UTI was correlated with other domains of quality, we assessed whether inappropriate diagnosis of UTI (as currently specified) was related to inappropriate diagnosis of CAP. This manuscript was published in *BMJ Quality & Safety*.[4] In it, we analyzed 10,398 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 UTI is moderately correlated with inappropriate diagnosis of CAP at the hospital level:

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




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.




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

**J. Empirical Validity: Association of Inappropriate diagnosis of UTI with Outcomes**

There are three main harms associated with inappropriate diagnosis of UTI: delayed time to true diagnosis, antibiotic-associated adverse events, and antibiotic resistance.

In a paper published in *JAMA Internal Medicine*, we analyzed outcomes associated with antibiotic treatment in 2,733 hospitalized patients with ASB (i.e., inappropriate diagnosis of UTI).[3] Patients inappropriately diagnosed with UTI were treated with a median (IQR) 7 (4-9) days of antibiotic therapy, all of which were unnecessary.

Outcomes of patients inappropriately diagnosed vs. those who had ASB and did not receive antibiotics are shown in the table below. Notably, patients inappropriately diagnosed with UTI who were treated with antibiotics had an approximately 1 day longer length of stay after date of urine testing than those who were not treated with antibiotics (aRR: 1.37 [1.28-1.47]).

Table 6. Outcomes for Treatment vs No Treatment for Asymptomatic Bacteriuria (N = 2733)

| Outcome[a] | Antibiotics (n=2259) | No Antibiotics (n=474) | Unadjusted Odds Ratio (95% CI) | Unadjusted *P* Value | Adjusted Odds Ratio (95% CI) | Adjusted *P* Value |
|---|---|---|---|---|---|---|
| 30-d Post discharge mortality[b], N (%) | 63 (2.8) | 11 (2.3) | 1.22 (0.66-2.26) | 0.53 | 1.34 (0.72-2.49) | 0.35 |

| Outcome[a] | Antibiotics (n=2259) | No Antibiotics (n=474) | Unadjusted Odds Ratio (95% CI) | Unadjusted *P* Value | Adjusted Odds Ratio (95% CI) | Adjusted *P* Value |
|---|---|---|---|---|---|---|
| 30-d Postdischarge readmission[b], N (%) | 362 (16.0) | 66 (13.9) | 1.16 (0.87-1.56) | 0.31 | 1.29 (0.92-1.81) | 0.14 |
| 30-d Post discharge ED Visit[b], N (%) | 272 (12.0) | 62 (13.1) | 0.91 (0.70-1.18) | 0.48 | 0.90 (0.66-1.24) | 0.52 |
| Discharge to post-acute care facility[b,c], N (%) | 811 (35.9) | 102 (21.5) | 1.98 (1.58-2.48) | <0.001 | 1.19 (0.90-1.57) | 0.22 |
| *Clostridioides difficile* infection[d], N (%) | 14 (0.6) | 2 (0.4) | 1.39 (0.41-4.68) | 0.59 | 0.88 (0.20-3.86) | 0.86 |
| Duration of hospitalization, median (IQR) d[e] | 4 (3-6) | 3 (2.5) | 1.37 (1.28-1.47)[f] | <0.001 | 1.37 (1.28-1.47)[f] | <0.001 |

Abbreviations: ED, emergency department; IQR, interquartile range.

[a] Outcomes were adjusted for patient variables found to be significant (P<.05) and associated with treatment in the bivariate and multivariate analysis.

[b] Mortality, readmissions, ED visits, and discharge to post-acute care facility were adjusted for age, Charlson Comorbidity Index score, hospitalization in 90 days preceding current admission, admission from nursing home, and insurance type.

[c] Long-term acute care hospital, skilled nursing facility, inpatient rehabilitation, and subacute rehabilitation.

[d] Infection occurring within 30 days of discharge was adjusted for age, history of antibiotic use and number of antibiotics in previous 90 days, admission from skilled nursing facility, prior hospitalization, proton-pump inhibitor use, immunosuppression, and Charlson Comorbidity Index score.

[e] From date of urine testing (either urine culture or urinalysis, whichever was performed first). Adjusted for age, sex, Charlson Comorbidity Index score, prior hospitalization, admission from nursing home, and insurance type.

[f] Relative risk given because duration of hospitalization is a continuous variable.

**[Response Ends]**

**2b.04. 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?)**

**[Response Begins]**

The validity of the inappropriate diagnosis of UTI 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 TEP members, patients, and end-users (hospitals); (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. As shown above, 82% of TEP members agreed with the statement: "The inappropriate diagnosis of UTI measure as specified can be used to distinguish between better and worse quality hospitals."

Perhaps even more importantly, 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 UTI 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), the low effect of abstraction errors on case classification, and by the long-standing general agreement by hospital experts with case classification during data feedback.

**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 UTI measure and measures of inappropriate diagnosis of other infections, namely CAP. As expected, we found hospitals that performed worse on one measure also performed worse on the other. Thus, the inappropriate diagnosis of UTI 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 UTI and outcomes.

**[Response Ends]**

**2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.**

*Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.*

**[Response Begins]**

We used the Spearman Brown prophecy to determine the minimum number of cases that hospitals participating in this measure would need to capture on an annual basis in order to allow us to distinguish performance accurately and reliably. Our analysis suggests that to meet the 0.8 standard for reliability, hospitals would need to abstract 59 cases annually.

Table 1. Number of annual cases needed to achieve each reliability threshold.

| Reliability | Number of annual cases needed |
|---|---|
| 0.6 | 22 |
| 0.7 | 35 |
| 0.8 (standard) | **59** |
| 0.9 | 132 |

Of the 40 hospitals participating in HMS in 2019 (our most recent year), 36/40 (90%) were able to meet this minimum standard of 59 annual cases (the 4 that did not were small hospitals). If we lowered the threshold for reliability to 0.7, 95% of hospitals would have been able to meet this minimum threshold of 35 cases.

To further characterize the degree of variability in the measure score we analyzed hospitals in the HMS cohort and:

1. Report the distribution of the measure score
2. Calculate the mean; standard deviation; median; and 10th, 25th, 75th, and 90th percentile of the performance scores for each quarter.

3. Group hospitals by quartiles and assess whether the difference in mean measure score between each adjacent quartile was statistically significant.

**[Response Ends]**

**2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.**

*Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.*

**[Response Begins]**

The distribution of the measure for all hospitals (each hospital=1 blue bar) is shown below in **Figure 1** with error bars representing 95% confidence intervals. **Table 2** shows summary statistics for all years combined, the first 4 quarters, and the final 4 quarters.

**Figure 1.** Distribution of Inappropriate Diagnosis of Urinary Tract Infection by Hospital

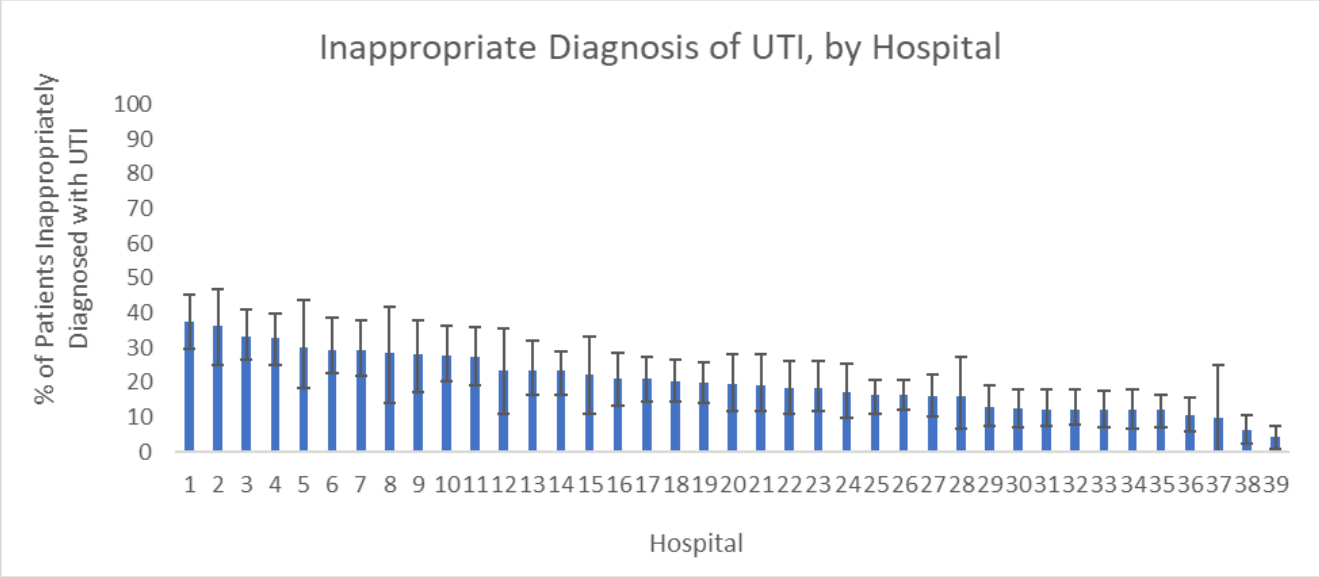



**Table 2.** Summary statistics for all years combined, the first 4 quarters, and the final 4 quarters.

| Year | Number of Hospitals | Number of UTI Patients | Overall Mean Inappropriate Diagnosis | Hospital Adjusted Mean (SD) | Min-Max | 10th Percentile (better performance) | 25th Percentile | Median | 75th Percentile | 90th Percentile (worse performance) |
|---|---|---|---|---|---|---|---|---|---|---|
| All years | 45 | 12,939 | 23.9% (3088/12939) | 24.7% (0.012) | 10.9%-47.4% | 14.7% | 18.9% | 23.2% | 30.6% | 38.0% |

| Year | Number of Hospitals | Number of UTI Patients | Overall Mean Inappropriate Diagnosis | Hospital Adjusted Mean (SD) | Min-Max | 10th Percentile (better performance) | 25th Percentile | Median | 75th Percentile | 90th Percentile (worse performance) |
|---|---|---|---|---|---|---|---|---|---|---|
| First 4 Quarters | 44 | 4,601 | 28.2% (1296/4601) | 28.3% (0.014) | 13.3%-53.5% | 16.7% | 2.0% | 26.8% | 33.7% | 43.1% |
| Last 4 quarters | 39 | 4,791 | 19.9% (954/4791) | 20.2% (0.013) | 4.2%-37.3% | 10.6% | 12.5% | 19.6% | 27.6% | 32.9% |

Compared with average-performing hospitals, hospitals in the 10th percentile (better performance) have about 12 fewer patients inappropriately diagnosed with UTI per 100 patients treated for UTI than the median (~84 fewer unnecessary antibiotic use days/100 UTI discharges), and hospitals in the 90th percentile (worse performing) 15 more patients were inappropriately diagnosed with UTI per 100 patients treated for UTI than the median (~105 more unnecessary antibiotic use days/100 UTI discharges).

The grouping of hospitals by quartiles for all years, first 4 quarters, and last 4 quarters, is shown in **Table 3**. All quartiles are statistically significantly different from other quartiles.

**Table 3.** Differences between adjacent quartiles of performance

| Percentile comparison | Lower Quartile | Higher Quartile | Test statistic | *p*-value |
|---|---|---|---|---|
| All years: 1st (best) quartile (0-25%) vs. 2nd quartile (25-50%) | 15.23% | 20.99% | 6.34 | <.001 |
| All years: 2nd (25%-50%) vs. 3rd quartile (50%-75%) | 20.99% | 26.80% | 5.30 | <.001 |
| All years: 3rd (50%-75%) vs. 4th (worst) quartile (75%-100%) | 26.80% | 36.03% | 7.36 | <.001 |
| First 4 quarters: 1st (best) quartile (0-25%) vs. 2nd quartile (25-50%) | 18.42% | 23.92% | 3.26 | 0.001 |
| First 4 quarters: 2nd (25%-50%) vs. 3rd quartile (50%-75%) | 23.92% | 30.92% | 3.56 | <.001 |
| First 4 quarters: 3rd (50%-75%) vs. 4th (worst) quartile (75%-100%) | 30.92% | 40.87% | 4.62 | <.001 |
| Last 4 quarters: 1st (best) quartile (0-25%) vs. 2nd quartile (25-50%) | 10.76% | 16.61% | 4.21 | <.001 |
| Last 4 quarters: 2nd (25%-50%) vs. 3rd quartile (50%-75%) | 16.61% | 22.47% | 3.50 | <.001 |
| Last 4 quarters: 3rd (50%-75%) vs. 4th (worst) quartile (75%-100%) | 22.47% | 31.75% | 4.95 | <.001 |

**[Response Ends]**

**2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.**

*In other words, what do the results mean in terms of statistical and meaningful differences?*

**[Response Begins]**

The measure was able to detect facilities with above- and below-average performance. In the first year, facility scores ranged from 13.3% to 53.5% with a mean performance of 28.3%. By the final year, facility scores had improved markedly and ranged from 4.2% to 37.3% with a mean performance of 20.2%.

Our analysis showed a statistically significant difference in performance between each quartile of hospitals, suggesting consistent performance gaps across facilities and targets for improvement.

**[Response Ends]**

**2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.**

*Describe the steps—do not just name a method; what statistical analysis was used.*

**[Response Begins]**

This measure is calculated using chart-abstracted data. To limit the effects of missing data, abstractors cannot submit a value of "missing" for individual data elements because the case will be rejected by the abstraction tool. Although abstractors cannot submit missing data, for some data (e.g., white blood cell count) they may submit a value of "unknown" or "not available." For cases submitted by hospitals from July 2017 through March 2020, we calculated the number of cases that were missing data used in case classification.

**[Response Ends]**

**2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.**

*For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).*

**[Response Begins]**

Data that were missing or marked as "unknown/not available" are presented below. Some of these data are accurately missing (e.g., no urinalysis obtained during hospitalization), others are missing due to errors.

As expected, missing data relevant to UTI cases were extremely rare. The percentage of encounters with missing, "unknown," or "not available" values was 5.2% (714/13,805) of all included patients.

Table 4. Percentage of encounters with missing, "unknown," or "not available" data

| Variable | Percent missing, "unknown," or "not available" % (n/N) |
|---|---|
| Age | 0% |
| Race | 1.8% (243/13,805) |
| Sex | 0% (6/13,805) |
| Ethnicity | 14.6% (2,019/13,805) |
| Temperature | 0% (4/13,805) |
| Heart rate | 0% (4/13,805) |
| Respiratory rate | 0% (4/13,805) |
| White blood cell count | 0.5% (73/13,805) |
| Urinary catheter | 3.5% (484/13,805) |
| Urine culture organism | 0% (5/13,805) |
| Urinalysis | 1% (144/13,805) |

**[Response Ends]**

**2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.**

*In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.*

**[Response Begins]**

The percentage of cases that could potentially be affected by missing data is negligible, indicating that missing data did not affect the performance results or other findings. As noted above, when data were missing it was often because they did not exist in the medical record (e.g., ethnicity), rather than due to an error in abstraction.

**[Response Ends]**

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

**2b.11. Indicate whether there is more than one set of specifications for this measure.**

**[Response Begins]**

No, there is only one set of specifications for this measure

**[Response Ends]**

**2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.**

*Describe the steps—do not just name a method. Indicate what statistical analysis was used.*

**[Response Begins]**

**[Response Ends]**

**2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.**

*Examples may include correlation, and/or rank order.*

**[Response Begins]**

**[Response Ends]**

**2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.**

*In other words, what do the results mean and what are the norms for the test conducted.*

**[Response Begins]**

**[Response Ends]**

**2b.15. Indicate whether the measure uses exclusions.**

**[Response Begins]**

Yes, the measure uses exclusions.

**[Response Ends]**

**2b.16. Describe the method of testing exclusions and what was tested.**

*Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?*

**[Response Begins]**

All exclusions were determined by careful clinical review and discussion and feedback from our national expert panel and HMS' Data, Design, and Publications Committee.

Exclusion criteria (and reasoning) include:

- Patients who left against medical advice or refused medical care
    - This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care

- Patients admitted on hospice or comfort care
    - This exclusion is needed for acceptability of the measure to hospitals, who may appropriately adjust their treatment and diagnostic procedures to comply with patient desires
- Patients who were pregnant or breastfeeding
    - This exclusion is needed for acceptability of the measure to hospitals, as pregnancy/breastfeed presents diagnostic and treatment challenges that may differ from patients who are not pregnant/breastfeeding
- Patients with a spinal cord injury
    - This exclusion was initiated by members of the TEP who believed this patient population to be substantially different from others included in the measures and to have potentially different signs and symptoms of a UTI. Thus, to increase acceptability and face validity, these patients are excluded.
- Patients with a UTI-related complication (operationalized by excluding patients discharged on more than 14 days of antibiotic therapy)
    - This exclusion is needed for acceptability of the measure to hospitals. UTI-related complications are not well documented on ICD or other coding but are important reasons to treat patients more aggressively. Generally, patients discharged on more than 14 days of antibiotics do not have typical UTIs; rather, they have an alternative reason or complication for extended therapy (e.g., nephric abscess).

To assess how common exclusion criteria were, we reviewed the literature—including national databases (Medicaid, Medicare, Premier) to estimate typical numbers of patients excluded for the above reasons. For the final exclusion criterion, we were able to estimate this directly from the HMS database.

**[Response Ends]**

**2b.17. Provide the statistical results from testing exclusions.**

*Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.*

**[Response Begins]**

Our exclusion results are shown below:

Table 1. Percent of individuals excluded based on exclusion criteria

| **Exclusion** | **Percent of Patients Excluded: Estimates from the Literature/HMS** |
|---|---|
| Patients who left against medical advice | 0.37% [1] |
| Patients who were pregnant or breastfeeding | 0.8% [2] |
| Patients admitted on hospice or comfort care | 0.33% (Medicaid) to 0.62% (Medicare)<br>1% (Premier) [3] |
| UTI-related complication<br>(>14 days of antibiotics at discharge) | 0.3% (9/3197)- HMS Estimates |
| **Total** | **1.8%-2.47%** |

In addition, we provided all exclusion criteria to participating hospitals and our technical expert panel to ensure they appeared feasible and reasonable. There was generally agreement across our groups that the exclusions led to a more accurate and fair assessment of patients inappropriately diagnosed with UTI. Spinal cord injury was one item discussed by the TEP who agreed they should be excluded. The TEP member from the American Urological Association reviewed our inclusion criteria for urinary anatomy and agreed with their operationalization. Some TEP members suggested additional populations to include in the future—such as surgical patients and those in nursing homes—but the group believed that

starting with a less contentious group (i.e., hospitalized medical patients) first would be a great start and a necessary step to move into more difficult populations (e.g., nursing homes).

[1] YNHHSC/CORE. Excess Days in Acute Care (EDAC) Measures Methodology. CMS.gov. Methodology Web site. https://qualitynet.cms.gov/inpatient/measures/edac/methodology. Published 2021. Accessed 11/20/2021.

[2] Dinh A, Ropers J, Duran C, et al. Discontinuing beta-lactam treatment after 3 days for patients with community-acquired pneumonia in non-critical care wards (PTC): a double-blind, randomised, placebo-controlled, non-inferiority trial. *Lancet*. 2021;397(10280):1195-1203. doi:10.1016/S0140-6736(21)00313-5.

[3] Lindenauer PK, Stefan MS, Shieh MS, Pekow PS, Rothberg MB, Hill NS. Outcomes associated with invasive and noninvasive ventilation among patients hospitalized with exacerbations of chronic obstructive pulmonary disease. *JAMA Intern Med*. 2014;174(12):1982-1993. doi:10.1001/jamainternmed.2014.5430. PCMID: PMC4501470.

**[Response Ends]**

**2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.**

*In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.*

**[Response Begins]**

Exclusions were uncommon. When present they were needed to improve acceptability by the hospitals. Feedback from our TEP and from end-user hospitals was supportive of the exclusions in their current form.

**[Response Ends]**

**2b.19. Check all methods used to address risk factors.**

**[Response Begins]**

No risk adjustment or stratification

**[Response Ends]**

**2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.**

**[Response Begins]**

**[Response Ends]**

**2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.**

**[Response Begins]**

N/A. Not an intermediate or health outcome, PRO-PM, or resource use measure.

In the context of healthcare performance assessment, the purpose of the risk model is to reduce bias due to case mix characteristics present at the start of care (i.e., to risk adjust), not to totally explain variation in outcomes, which would

require also including variables about quality of care. Variables related to quality of care are purposely not included in risk models for performance measures used to assess quality.[4]

Specifically, CMS notes:

- "Process measures are not risk-adjusted; rather the target population of a process measure is defined to include all patients for whom the process measure is appropriate."
- "The variation in measured entity-level (e.g., clinician or facility) performance may be due to variation in quality or variation in factors that are independent of quality (e.g., factors like the age or severity of illness of patients). Independent of quality means that the clinician treats the patients exactly the same way, but patients who have the factor (older or sicker) have worse outcomes than patients who do not (younger or less sick)."

[4] Measures Management System Risk Adjustment. Centers for Medicare & Medicaid. Measure Management & You Web site. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Risk-Adjustment.pdf. Published 2017. Accessed 11/30/2021.

**[Response Ends]**

**2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.**

**[Response Begins]**

**[Response Ends]**

**2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.**

*Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).*

**[Response Begins]**

**[Response Ends]**

**2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.**

**[Response Begins]**

**[Response Ends]**

**2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.**

*Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.*

**[Response Begins]**

**[Response Ends]**

**2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.**

*Validation testing should be conducted in a data set that is separate from the one used to develop the model.*

**[Response Begins]**

**[Response Ends]**

**2b.27. Provide risk model discrimination statistics.**

*For example, provide c-statistics or R-squared values.*

**[Response Begins]**

**[Response Ends]**

**2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).**

**[Response Begins]**

N/A. Not an intermediate or health outcome, PRO-PM, or resource use measure. No risk model/stratification.

**[Response Ends]**

**2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.**

*The preferred file format is .png, but most image formats are acceptable.*

**[Response Begins]**

**[Response Ends]**

**2b.30. Provide the results of the risk stratification analysis.**

**[Response Begins]**

**[Response Ends]**

**2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).**

*In other words, what do the results mean and what are the norms for the test conducted?*

**[Response Begins]**

**[Response Ends]**

**2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.**

*Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.*

**[Response Begins]**

**[Response Ends]**

## Criteria 3: Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

---

**3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.**

**[Response Begins]**

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

**[Response Ends]**

**3.02. Detail to what extent the specified data elements are available electronically in defined fields.**

*In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.*

**[Response Begins]**

Some data elements are in defined fields in electronic sources

**[Response Ends]**

**3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.**

**[Response Begins]**

Currently, some of the inappropriate diagnosis of UTI data elements can be captured electronically in discrete fields (e.g., vital signs, laboratory values). However, not all documentation required to report the inappropriate diagnosis of UTI measure can be captured electronically in discrete fields. In particular **symptoms of UTI** are not in defined, computer-readable fields.

**Rationale for Using Data Elements not from Electronic Sources**

While efforts are being made to facilitate an electronic measure (see below), gaps remain in the ability to electronically capture all of the required data for measure validity. The inappropriate diagnosis of UTI measure requires data abstractors to review documentation in various formats, including narrative free text, to identify the specific information necessary to report the measure. Preliminary efforts to convert the inappropriate diagnosis of UTI measure to an eCQM within the current Health Quality Measure Format/Quality Data Model frameworks showed that the transition is not immediately feasible.

Symptoms are generally documented in free-text spaces within the medical record, and their location varies by hospital. Symptoms are critical to measure validity, as urine cultures and other diagnostic tests (e.g., urinalyses, white blood cell counts) are both insensitive and non-specific, and UTI is a clinical diagnosis.[1-3] Measures of diagnostic accuracy of UTI thus require clinical data—namely, symptoms.

**Possible Replacement for Free-text Symptoms**

Because symptoms are the primary reason the measure cannot be an eCQM, we tested a method that replaces **free text symptoms** with a discrete data element—**urine culture indications**. Some (but not all) hospitals allow or require an indication when ordering a urine culture. Indications can be free text but are often listed from discrete variables that include symptoms or signs of infection. Assuming the listed signs or symptoms in the indication are accurate (i.e., the clinician is selecting an accurate choice), then they could feasibly be used instead of free text symptoms from the medical record. We tested whether this method would be valid in the HMS cohort of hospitals.

First, we found that over half of patients had no indication listed in the urine culture order, including 61.9% (1979/3197) of inappropriate diagnosis of UTI cases and 56.4% (5981/10,608) of UTI cases. Another quarter of patients (27.4% of inappropriate diagnosis and 24.7% of UTI cases) were the result of "reflex" cultures triggered by "positive" urinalyses and thus did not have an indication listed. This left approximately 25% of all patients with an indication listed. Of those, the most common indication listed was abnormal urinalysis; see the following table for all listed indications:

Table 1. Top 10 Indications for Urine Cultures (Indications Listed in the Urine Culture Order), N=3,981

| Discrete Urine Culture Indication | UTI by Free Text Symptoms, N=3,167 | Inappropriate Diagnosis of UTI by Free Text Symptoms, N=814 |
|---|---|---|
| Abnormal Urinalysis | 53% | 68% |
| Other | 13% | 14% |
| Dysuria | 13% | 5% |
| Altered Mental Status | 11% | 10% |
| Fever | 5% | 1% |
| Frequency | 5% | 1% |
| Costovertebral Angle Pain/Tenderness | 5% | 1% |
| Hematuria | 4% | 0% |
| Suprapubic Pain | 4% | 1% |
| Urgency | 3% | 1% |
| Abdominal Pain | 3% | 2% |

*May add up to more than 100% as patients could have multiple indications.

After excluding cases with no indication in the order (n = 7,960), cases called UTI due to presence of severe sepsis or bacteremia (n = 1,430), and cases that were urine reflex cultures (n = 3,494), we compared the classification of cases as inappropriate diagnosis of UTI vs. UTI by urine culture indication to classification by chart review to determine sensitivity, specificity, negative predictive value, and positive predictive value.

We found that discrete urine culture indications have a high sensitivity but low specificity for identifying inappropriate diagnosis of UTI. This indicates there is a low positive predictive value and high negative predictive value for identifying inappropriate diagnosis of UTI. Thus, if the indication in the urine culture order indicates inappropriate diagnosis of UTI, the order is impossible to interpret. If the indication in the urine culture order indicates UTI, the order is likely to be correct – in part because UTI is more common.

Table 2. Case Classification by Chart Review vs. Urine Culture Indication, N=3150

| * | UTI by Free Text Symptoms, N=2,438 | Inappropriate Diagnosis of UTI by Free Text Symptoms, N=712 |
|---|---|---|
| **UTI by Discrete Urine Culture Indication, N=1,301** | True UTI<br>1,238 | False UTI<br>63 |
| **Inappropriate Diagnosis of UTI by Discrete Urine Culture Indication, N=1,849** | False Inappropriate Diagnosis of UTI<br>1,200 | True Inappropriate Diagnosis of UTI<br>649 |

*Indicates cell intentionally left blank

Sensitivity of discrete urine culture indication for Inappropriate Diagnosis of UTI =

True Inappropriate Diagnosis of UTI/(True Inappropriate Diagnosis of UTI + False UTI) =

**91.2%** Specificity of discrete urine culture indication for Inappropriate Diagnosis of UTI =

True UTI/(True UTI + False Inappropriate Diagnosis of UTI) = **50.8%**

Positive Predictive Value (for identifying Inappropriate Diagnosis of UTI) =

True Inappropriate Diagnosis of UTI/(True Inappropriate Diagnosis of UTI/False Inappropriate Diagnosis of UTI) = **35.1%**

Negative Predictive Value (for identifying UTI) = True UTI/(True UTI + False UTI) = **95.2%**

Based on this analysis, at this time, urine culture indications are not a valid way of capturing symptoms and classifying patients as overdiagnosis of UTI vs. UTI. This is consistent with prior studies of catheter-associated UTI which found little overlap between symptoms noted in discrete and free text fields.[4] However, urine culture indications are a credible, near-term path to eCQM. For that to happen, hospitals would need to expand efforts to require urine culture indications (a process already underway), and indication accuracy would need to improve. Alternatively natural language processing could be developed to identify symptoms from the free-text areas of the medical record (see below).


1. Choudhuri JA, Pergamit RF, Chan JD, et al. An Electronic Catheter-Associated Urinary Tract Infection Surveillance Tool. Infection Control & Hospital Epidemiology. 2011;32(8):757-762.
2. Branch-Elliman W, Strymish J, Kudesia V, Rosen AK, Gupta K. Natural Language Processing for Real-Time Catheter-Associated Urinary Tract Infection Surveillance: Results of a Pilot Implementation Trial. Infect Control Hosp Epidemiol. 2015;36(9):1004-1010.
3. Wald HL, Bandle B, Richard AA, Min SJ, Capezuti E. Implementation of electronic surveillance of catheter use and catheter-associated urinary tract infection at Nurses Improving Care for Healthsystem Elders (NICHE) hospitals. Am J Infect Control. 2014;42(10 Suppl):S242-249.
4. Sanger PC, Granich M, Olsen-Scribner R, et al. Electronic Surveillance For Catheter-Associated Urinary Tract Infection Using Natural Language Processing. AMIA Annu Symp Proc. 2017;2017:1507-1516.


**[Response Ends]**

**3.04. Describe any efforts to develop an eCQM.**

**[Response Begins]**

Fortunately, as noted above, there are multiple promising pathways to eCQM development. First, there are methods under development to assess diagnostic accuracy of UTI using natural language processing to identify symptoms from the medical record. For example, one single-center study at the University of Washington was able to identify catheter-associated UTI vs. catheter-associated asymptomatic bacteriuria using natural language processing to identify symptoms.[5] Another study tested natural language processing to identify urinary symptoms in hospitalized patients and found high sensitivity (100%) and positive predictive value (97%).[1] These methods need to be prospectively validated in different settings and for non-catheter-associated UTI; however, they show promise for eCQM development. Second, requiring urine culture indications is becoming a more standard process for hospitals which is likely to improve the sensitivity and specificity of indications for identifying overdiagnosis of UTI. Multiple research efforts (e.g., through the Centers for Disease Control and Prevention Shepherd projects and through the Gordon and Betty Moore Foundation) continue to make progress on eCQM development for UTI diagnosis.


1. Gundlapalli AV, Divita G, Redd A, et al. Detecting the presence of an indwelling urinary catheter and urinary symptoms in hospitalized patients using natural language processing. J Biomed Inform. 2017;71s:S39-s45.


**[Response Ends]**

**3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

**[Response Begins]**

**Data Collection, Availability, Missing Data**

This measure is calculated using chart-abstracted data routinely collected during the normal process of patient care and requires no additional data. Because these data are captured as standard practice, missing data were extremely rare. The percentage of encounters with any missing, "unknown," or "not available" values was 5.2% (714/13,805) of all included patients. This missing data had little effect on the ability to classify the case as inappropriate vs. appropriate diagnosis of UTI.

**Timing/Frequency of Data Collection and Patient Sampling**

Hospitals have the option to sample from their population or submit their entire population. Hospitals also have the option to sample quarterly or monthly. Over the entire year, 59 cases are recommended for the denominator. Thus, hospitals whose patient population size is less than or equal to the minimum number of cases per quarter (N=15) or month (N=5) for the measure should not sample. A hospital may choose to use a larger sample size than is required.

Using the current HMS hospital cohort as a representative example, the minimum number of case abstracts per hospital per year to meet pre-specified reliability thresholds of 0.7 and 0.8 are highly attainable. Within a cohort of 40 HMS hospitals participating in 2019, 90% of hospitals were able to abstract the minimum of 59 cases to achieve 0.8 reliability. Of those that could not abstract the required number of cases, hospital bed sizes were 49 beds, 68 beds, 75 beds, and 133 beds. Ninety-five percent of hospitals could abstract the 35 cases/year necessary to achieve 0.7 reliability, and all but one could reach the abstraction threshold for 0.6 reliability. Of the two hospitals unable to achieve abstraction thresholds for 0.7 reliability (75 beds and 133 beds), one hospital over-sampled cases for an alternative measure and the other had challenges with data abstractor hiring.

**Patient Confidentiality**

Data are de-identified.

**Time and Cost**

To improve feasibility and reduce time and cost of data collection, we removed all non-essential data collection elements from the measure during measure testing. We also reviewed exclusion criteria to remove those that were uncommon and would not impact measure outcomes. This pared down data collection form was tested at 4 hospitals in Utah to estimate the time needed for case review. Those results follow:

- Review of eligibility criteria to determine whether a patient would be included vs. excluded took 1-3 minutes.
- Review time could be reduced by adding exclusion criteria (e.g., ICU admission) electronically to lists for review.
- Across the 4 hospitals, 34.7%-69.3% of patients reviewed for inclusion were eligible (see Table for details)
- Once determined to be eligible, case review took 15 to 30 minutes per case.

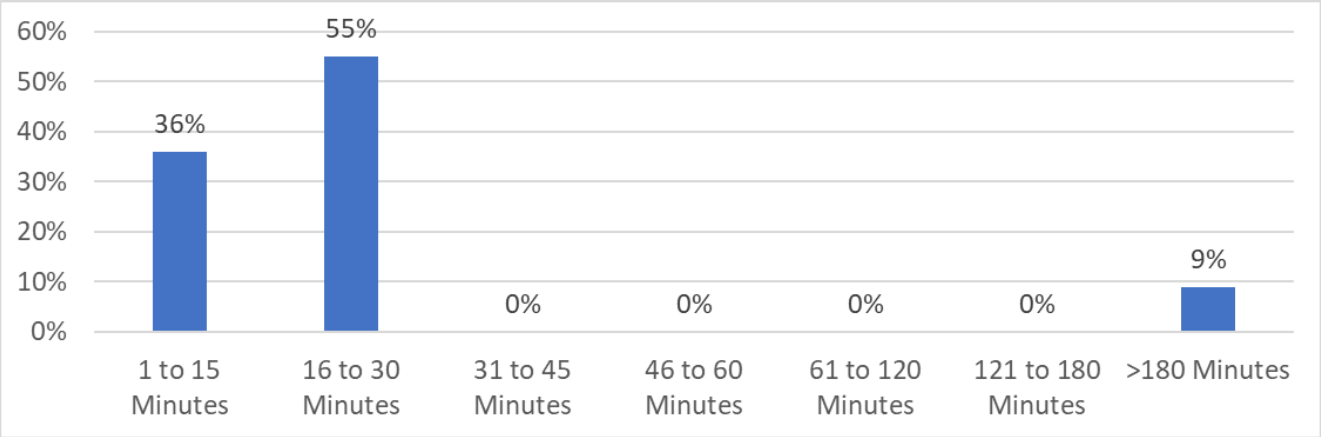
Table 3. Case review of hospitalized patients with positive urine culture at 4 Utah hospitals over a 6-month period.

| * | # beds | # cases reviewed | # cases included | % included |
|---|---|---|---|---|
| **Hospital 1** | 1000 | 216 | 75 | 34.7% |
| **Hospital 2** | 502 | 75 | 52 | 69.3% |
| **Hospital 3** | 90 | 75 | 38 | 50.6% |
| **Hospital 4** | 132 | 84 | 52 | 61.9% |

*Indicates cell intentionally left blank

When speaking to Infection Preventionists at included hospitals, the time for data collection was on par with other NHSN measures currently requiring case review (e.g., CAUTI, CLABSI, SSI, CDI, VAP). They all noted that feasibility improved for those measures over the years as electronic health record vendors built modules to reduce initial screening. The Joint Commission also provided comparative data during our Technical Expert Panel. They noted that 4 chart review measures abstracted across 11 sites had similar time requirements to our proposed measure.

**Time Required for Abstraction of 4 Different Measures**




*Data provided by Dr. David Baker of The Joint Commission

During our technical expert panel, we surveyed our experts on measure feasibility via the following two questions:

1. **How appropriate is the quantity of information collected for use in determining inappropriate diagnosis of UTI? (N=11 experts)**

- 82% (9/11) responded it was the correct amount of data

1. **Compared to other measures requiring chart review, how easy do you believe it would be for a hospital to collect the data needed to assess whether a case represents an inappropriate diagnosis of UTI? (N=11 experts)**

- 46% (5/11) reported it would be "about the same as other measures"
- 27% reported it would be easier and 27% reported it would be more difficult than other measures

We also surveyed hospitals participating in HMS (N=40) to ask about their experiences with the feasibility of the inappropriate diagnosis of UTI measure (see Table 4).

Table 4. Responses to the question: **How easy is it for your hospital to collect the data needed to assess whether a case represents asymptomatic bacteriuria? (N=40 hospitals)**

| N=40 hospitals | Response; N (%) |
|---|---|
| Very Easy | 4 (10.0%) |
| Easy | 11 (27.5%) |
| Neither Easy nor Difficult | 16 (40%) |
| Difficult | 7 (17.5%) |
| Very Difficult | 2 (5%) |

The majority of respondents reported it was very easy, easy, or neither easy nor difficult: 31/40 (77.5%)

**[Response Ends]**

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

**3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),**

**Attach the fee schedule here, if applicable.**

**[Response Begins]**

All measures are free to use. Data dictionaries and data collection templates are free and accessible at our website (https://mi-hms.org/inappropriate-diagnosis-urinary-tract-infection-uti-hospitalized-medical-patients ).

**[Response Ends]**

## Criteria 4: Use and Usability

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

---

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

**4a.01.**

**Check all current uses. For each current use checked, please provide:**

**Name of program and sponsor**

**URL**

**Purpose**

**Geographic area and number and percentage of accountable entities and patients included**

**Level of measurement and setting**

**[Response Begins]**

Payment Program

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Program: Michigan Hospital Medicine Safety Consortium (HMS)

Sponsor: Blue Cross and Blue Shield of Michigan

URL:

https://mi-hms.org/quality-initiatives/antimicrobial-use-initiative

Purpose: To improve outcomes of hospitalized patients treated for urinary tract infection (UTI).

Geographic area: Acute care hospitals in the state of Michigan. Between 7/1/2017 and 3/31/2020 there were 13,805 hospitalized patients treated for UTI across 49 HMS hospitals.

Level of measurement and setting: We collect patient-level data which is evaluated for inappropriate diagnosis of UTI. Since January 1, 2018, HMS hospitals have received financial incentives based on their performance on the inappropriate diagnosis of UTI measure. Annual target goals are established by the HMS Coordinating Center and approved by the HMS Data, Design, and Publications Committee and the funder (Blue Cross and Blue Shield of Michigan). Goals are meant to be "stretch" goals that drive hospitals to improve every year.

Program: Michigan Hospital Medicine Safety Consortium (HMS)

Sponsor: Blue Cross and Blue Shield of Michigan

URL:

https://mi-hms.org/quality-initiatives/antimicrobial-use-initiative

Purpose: To improve outcomes of hospitalized patients treated for urinary tract infection (UTI).

Geographic area: Acute care hospitals in the state of Michigan. Between 7/1/2017 and 3/31/2020 there were 13,805 hospitalized patients treated for UTI across 49 HMS hospitals.

Level of measurement and setting: We collect patient-level data which is evaluated for inappropriate diagnosis of UTI. Hospitals receive a list of all patients considered inappropriately diagnosed. In addition, aggregated data on inappropriate diagnosis of UTI from each hospital is presented quarterly and annually to hospitals to allow them to compare: a) performance in their own hospital over time and b) performance compared to other hospitals participating in HMS.

**[Response Ends]**

**4a.02. Check all planned uses.**

**[Response Begins]**

Public reporting

Public Health/Disease Surveillance

Payment Program

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Quality Improvement (internal to the specific organization)

Measure Currently in Use

**[Response Ends]**

**4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.**

*For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?*

**[Response Begins]**

**[Response Ends]**

**4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.**

*A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*

**[Response Begins]**

**[Response Ends]**

**4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

*Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.*

**[Response Begins]**

Since 2017, the inappropriate diagnosis of UTI measure has been in use through the Michigan Hospital Medicine Safety Consortium (HMS) to measure and improve care for hospitalized patients with UTI. HMS is a collaborative quality initiative of 60+ hospitals across the state of Michigan whose purpose is to improve the care of hospitalized infections. As part of its Antimicrobial Use Initiative, data have been collected from a pseudo-random population of hospitalized patients treated for UTI. Every quarter, participating hospitals receive data on the proportion of patients treated for UTI at their hospital that are inappropriately diagnosed. In addition, each hospital receives data on how their performance compares to all other hospitals in HMS and how their performance has changed over time. Hospitals also receive a list of patients who were considered inappropriately diagnosed so that they can further evaluate inappropriate diagnosis and use those data to drive internal quality improvement efforts.

Beginning in 2018, a pay-for-performance incentives was initiated for HMS hospitals whereby a percentage of their Blue Cross and Blue Shield of Michigan reimbursements were given if they met a pre-defined performance metric. Every year since, the threshold for full payment has been made harder in order to continue to drive improvement.

**[Response Ends]**

**4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

**[Response Begins]**

***Tri-annual Collaborative Wide Meetings***

Individuals from participating hospitals meet in person three times a year. We encourage hospitals to send their Clinical Data Abstractors, physician champions, and quality leads, as well as other individuals from their hospital that might be interested in participation. These meetings take place three times per year – in March, July, and November. Traditionally, meetings took place in-person at venues across Michigan. In 2020 and 2021, these meetings were hosted via an on-line format due to COVID-19.

The tri-annual meetings provide individuals from member hospitals with the opportunity engage with each other in a variety of formats. Each meeting includes a formal discussion of the data from each of the HMS initiatives — including data on inappropriate diagnosis of UTI — for the previous quarter, presentations from member hospitals and expert guests, breakout/work group sessions, and networking opportunities. These meetings allow individuals from member hospitals to network with individuals from other hospitals who have excelled in those areas to seek ideas on how to improve their performance. It also allows for an opportunity for feedback and to answer questions related to their measure performance.

***Site-specific Reports on Measure Performance***

Tri-annually, each participating hospital receives a printed and email version of a site-specific data report. These reports are also available daily within the database/registry (see below). These reports provide an in-depth look into the performance of each site. For example, we provide hospital data on the number of patients inappropriately vs. appropriately diagnosed with UTI, details on antibiotic use and outcomes (e.g., adverse events), longitudinal performance, and data on how individual hospitals compare to other hospitals in the state in terms of inappropriate diagnosis. Hospitals also receive a list of all patients who were considered "inappropriately diagnosed with UTI" to enable them to return to their hospital and conduct case reviews of those patients. Each hospital is encouraged to review these cases with their local team to perform audit and feedback, identify trends, and assist with overall quality improvement. This also provides an opportunity for measure feedback — for example, hospitals might find an error in case classification. Early during measure development this case-specific feedback was critical for improving measure validity.

***Live Database Reports***

Each of the HMS databases are equipped with the ability to view live reports utilizing Business Objects software. These reports provide updated data every 24 hours regarding measures (site performance and collaborative performance), fallout case information, demographics, critical/non-critical data errors, completeness of abstracted cases, and case classification information.

Individuals who participate in the collaborative either as a Clinical Data Abstractor or a quality administrator have the ability to log into the HMS databases and view these reports at their leisure. The software that HMS utilizes also allows for these reports to be exported as Excel files or PDFs for hospital-specific customization. This information is often utilized by participating hospitals at committee meetings or for presentations to track progress and inform quality improvement efforts. They also assist the Clinical Data Abstractor to identify errors in their abstraction and resolve them in real time. These reports also allow hospitals to review individual fallout cases and their clinical scenarios to inform individual clinicians or groups of clinicians of their performance and provide targeted education.

**[Response Ends]**

**4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.**

**[Response Begins]**

Throughout measure development, we received feedback on the measure performance/validity through three mechanisms: 1) Expert Feedback from Data Design and Publications Committee and Michigan Hospital Medicine Safety (HMS) Consortium Hospital Experts/Representatives, 2) "Fall-out" Feedback, and 3) October 2021 Hospital Survey.

Feedback from the Data, Design, and Publications Committee and "Fall-out" feedback has been described in the "validity" section. Briefly, measure performance feedback allowed us to refine the measures to the current version. The Data, Design, and Publications Committee approved the measures for use across HMS.

In October 2021, we systematically assessed the perceived use and usability of the inappropriate diagnosis of UTI measure by soliciting feedback from HMS hospitals participating at that time (N=40) via an online survey. Specifically, we asked all participating hospitals (N=40) to answer the following questions:

> Q1. Please briefly describe how you have used or are planning to use the [inappropriate diagnosis of UTI] measure to improve care.

**Responses:** The 40 responses to this open-ended question largely fell into a few broad categories. The majority of hospitals are using strategies related to audit, feedback and education. Examples include "have used it to provide feedback to clinicians in cases of inappropriate use, as one more tool discouraging antibiotic use"; "present data to physicians, review ASB fallouts with ED physicians"; or "we discuss the measure with providers, especially when discussing fallouts, and then asking what or if we could have done anything differently". There are a few hospitals that are using this data to update their tools or order sets. For example, "we have used numbers to modify ordering reflect UA, removing urine cultures from order set" or "revising clinical decision support tools".

> Q2. What perceived barriers do you see/foresee to using the [inappropriate diagnosis of UTI] measure to guide care improvement?

**Responses:** One-third of hospitals (45.0%, 14/40) of hospitals indicated that they don't see/foresee any barriers. Another third (30.0%, 12/40) noted issues with physician pushback/buy in. Statements made here include "physician resistance – change is difficult for many" or "physicians continuing to prescribe antibiotics based on old practices". There was also a broad category that related specifically to the treatment of patients who were confused or had "altered mental status" for a UTI (15.0%, 6/40), which includes "continued misinformation about elderly patients confused equates to UTI," or "patients with dementia or other clinical conditions who present to the ED, especially those who have a history of UTI in the past are often cultured despite having no symptoms". Finally, several participants noted challenges with education

and feedback, such as "The time required to educate and provide feedback to providers on how they are meeting the ASB measure goals."

**[Response Ends]**

**4a.08. Summarize the feedback obtained from those being measured.**

**[Response Begins]**

In summary, feedback from hospitals on measure performance was used to inform the development and refinement of the measures as currently submitted. In addition, feedback on measure implementation was broadly positive —that the measures were useful to guide care and improve diagnosis and antibiotic use. Based on feedback that time was a barrier to data collection, we limited the amount of data to be collected (average time for case collection 15-30 minutes) and decreased the number of cases we request be abstracted to still achieve a high reliability (N=59 cases). Thus, the measure submitted in this proposal should have even higher feasibility with similar usability as the measure tested in the Michigan Hospital Medicine Safety Consortium.

**[Response Ends]**

**4a.09. Summarize the feedback obtained from other users.**

**[Response Begins]**

In addition to the hospital feedback described above, we conducted a Patient Engagement Panel in order to understand patient perspectives about antibiotic treatment during hospitalization with infection. Seven individuals who were hospitalized or had a close family member who was hospitalized for an infection and received antibiotics participated in a 90-minute focus group. A discussion guide was used to assess participants' knowledge and perceptions about: how diagnoses are made and what information is needed; antibiotic risks and benefits; certainty of diagnosis and timing of treatment initiation; whether knowing how well a hospital accurately diagnoses infections would influence treatment choices. A brief summary of the Patient Engagement Panel is presented below.

Table 1. Summary of Patient Engagement Panel

| Question/Topic | Responses | Impression |
|---|---|---|
| Understanding of how infection diagnosis was made | Patients were aware of the necessity of tests (e.g., chest X-rays), labs (e.g., urine and blood tests), and clinical signs and symptoms (e.g., fever, $O_2$ saturation, pain, cough) in determining the diagnosis of infection. They relied on physicians' knowledge, but in some instances understood there may disagreement. | Patients understood that a process is involved in diagnosis; that diagnosis is reliant on lab results (which take time); and that there may be some uncertainty and thus differing opinions of physicians. |
| Risks and Benefits of Antibiotics | Patients universally agreed that antibiotics are beneficial: quickly reducing symptoms and clearing infections; necessary for treatment of severe illness. The discussion of risks identified many concerns: antibiotic resistance, allergic reactions, disruptions to gut microbiome, side effects from drug:drug interactions. | Patients understood there were both benefits and risks of antibiotic treatment. |

| Question/Topic | Responses | Impression |
|---|---|---|
| What does over-diagnosis mean? Under-diagnosis? | Patients expressed several ideas about what "over-diagnosis" is: "prescribing medication whether needed or not", "when a minor issue is overemphasized and overtreated", "antibiotics given without tests being done". The idea of "under-diagnosis" was expressed as: "settling on a routine diagnosis when something more significant is happening", "not utilizing antibiotics", "not enough concern when treating a routine" infection. | Patients understood that "over-diagnosis" relates to treatment that may not be necessary and that "under-diagnosis" involves the possibility of missing the diagnosis and not receiving the appropriate treatment. |
| How do you know if a hospital is doing a good job? What would help you to know? | Patients were aware that hospitals are rated on certain performance measures. They also expressed some skepticism about these due to: not knowing what the ratings are based on, variations in individual physicians (e.g., a top-rated hospital could still have a low-rated physician and vice versa), concern that hospitals could "game" the system of measurement. Even so, patients expressed interest in being able to access ratings of performance for aspects of healthcare. | Patients were receptive to information about hospital performance measures, especially if they had some assurances that they could be trusted. They were interested in measures of diagnostic performance as a way to make informed decisions about hospitals. |

Summary of patient feedback: Based on the focus group discussion, the measure is consistent with their understanding and expectations of diagnosis and treatment of infection. There were no issues or concerns raised that would necessitate modifications of the measure.

**[Response Ends]**

**4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

**[Response Begins]**

Feedback from HMS hospitals, the technical expert panel, and the patient engagement panel was all used to refine the measure. Major changes include: a) simplification of measure, b) refinement of measure specifications, c) streamline/decrease in amount of data requested for assessment, and d) defining minimum cases necessary for abstraction to decrease number of cases required to be submitted. We also received feedback on the naming of the measure. When we first began measure development, the measure had been named "over-diagnosis of UTI" which we changed to "inappropriate diagnosis of UTI" based on feedback from diagnostic error experts in our technical expert panel and to avoid confusion as "over-diagnosis" has alternate meanings in the diagnostic error community.

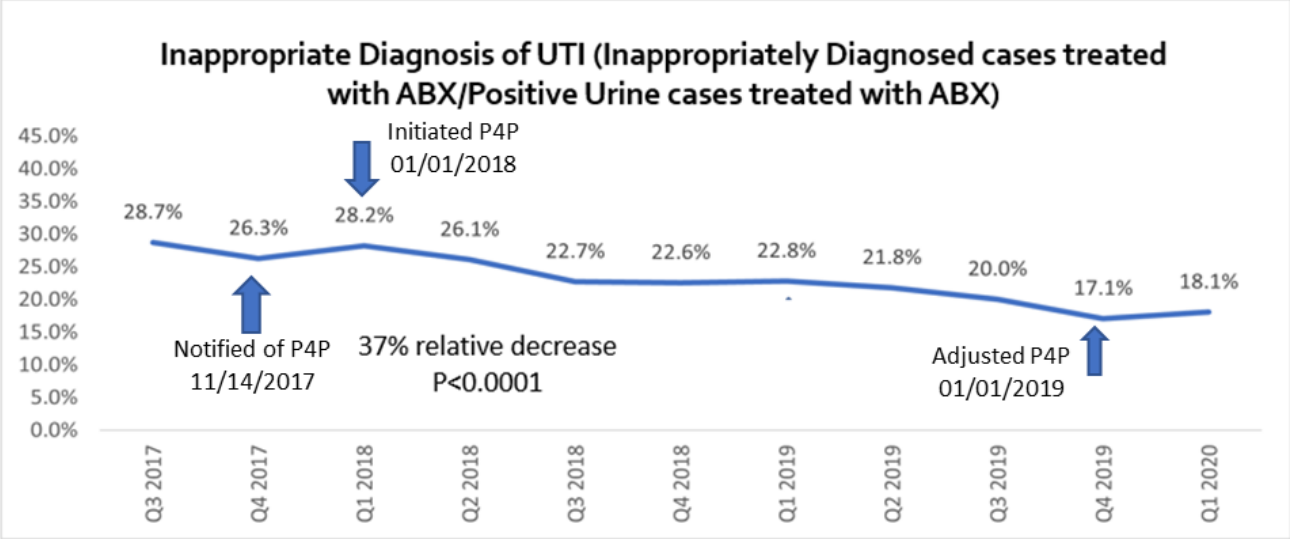
**[Response Ends]**

**4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement**

**at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

**[Response Begins]**

Since 2017, when the measure first began being reported to 49 participating HMS hospitals, we have seen a 37% relative decrease (P<0.001) in the percentage of patients inappropriately diagnosed out of all patients treated for UTI (see Figure). This represents an improvement in diagnosis, reduction in unnecessary antibiotic use, and improved care. The arrows show times when HMS pay-for-performance measures were announced, initiated, and the adjusted to continuously drive improvement.




In addition, since 2017, we have seen a statistically significant (though minor) decrease in antibiotic duration for patients inappropriately diagnosed with UTI (driven mostly by fewer cases of excess duration longer than 8 days). Though no antibiotic therapy is ideal for this patient population, there is often diagnostic uncertainty that drives brief empiric therapy. Stopping this therapy as soon as possible can reduce the risk of harm. In fact, in a paper published in JAMA Internal Medicine, we found that unnecessary antibiotic use for patients inappropriately diagnosed with UTI is associated with a longer hospital length of stay (adjusted relative risk 1.33 [1.22-1.46] or ~1 day longer length of stay).[1]


1. Petty LA, Vaughn VM, Flanders SA, et al. Risk Factors and Outcomes Associated With Treatment of Asymptomatic Bacteriuria in Hospitalized Patients. *JAMA Intern Med.* 2019;179(11):1519–1527. doi:10.1001/jamainternmed.2019.2871


**[Response Ends]**

**4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.**

**[Response Begins]**

There were no unexpected findings. Expected findings included decreased rates of inappropriate diagnosis of UTI, decreased unnecessary antibiotic use, and decreased length of stay.

In October 2021, we systematically assessed the perceived use and usability of the inappropriate diagnosis of UTI measure by soliciting feedback from HMS hospitals participating at that time (N=40). Via online survey, we asked all hospitals to answer the following questions:

1. What unintended consequences do you see/foresee to using the [inappropriate diagnosis of UTI] measure to guide care improvement? (Q547)

Over half of respondents said none/unknown (24/40; ). One-quarter (n=10) noted lack of appropriate treatment, including delays in treatment or missing alternative diagnoses. For example, "may be missing diagnosis of acute infection" or "possible delays in antibiotics in patient who actually require them". Other respondents noted issues related to dementia/altered mental status or patient dissatisfaction if they do not receive antibiotics.

**[Response Ends]**

**4b.03. Explain any unexpected benefits realized from implementation of this measure.**

**[Response Begins]**

Generally, there were no "unexpected benefits." Expected benefits included decreased rates of inappropriate diagnosis of UTI, decreased unnecessary antibiotic use, and decreased length of stay.

In October 2021, we systematically assessed the perceived use and usability of the inappropriate diagnosis of UTI measure by soliciting feedback from HMS hospitals participating at that time (N=40). Via online survey, we asked all hospitals to answer the following question:

1. If you have already started work based on the [inappropriate diagnosis of UTI] measure, what unexpected benefits have been realized from implementing this measure?

   **Responses**: 6: N/A, 5: none, 3: unsure/not sure; 2: question not answered; 1: Early stages of project

   A number of respondents (12) identified improved antibiotic use, either in terms of fewer patients treated inappropriately or through a reduction of complications due to antibiotic overuse. Five individuals noted a culture of less testing, particularly as it relates to urine cultures. Three individuals identified increased awareness in general around treatment of UTI and asymptomatic bacteriuria. Other responses included improved patient care, and alignment of system-wide culturing.

**[Response Ends]**

## Criteria 5: Related and Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

---

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

**5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).**

*(Can search and select measures.)*
**[Response Begins]**
0684: Percent of Residents with a Urinary Tract Infection (Long Stay)
0138: National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure
**[Response Ends]**

**5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).**

*(Can search and select measures.)*

**[Response Begins]**
**[Response Ends]**

**5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.**

**[Response Begins]**
N/A
**[Response Ends]**

**5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.**

**[Response Begins]**
No
**[Response Ends]**

**5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

**[Response Begins]**
NQF 0138, National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure, provides the Standardized Infection Ratio (SIR) of healthcare-associated CAUTIs. The target population, patients with chronic catheter use, is a subset of the target population for the Inappropriate Diagnosis of UTI measure. The focus of the NQF 0138 measure is primarily to prevent CAUTI by reducing foley catheter use and improving insertion practices. Our measure addresses inappropriate treatment of patients with antibiotics when they do not actually have UTI or CAUTI. Thus, rather than preventing CAUTI, our measure is focused on preventing an inappropriate diagnosis of CAUTI and subsequent antibiotic use. The measures include overlapping populations but have different goals and outcomes. NQF 0684, Percent of Residents with a Urinary Tract Infection (Long Stay), reports the percentage of long-stay nursing home residents who have a urinary tract infection the 30 days prior to assessment, based on data from the Minimum Data Set (MDS) 3.0 OBRA, PPS, and/or discharge assessments during the selected quarter for the purpose of reducing UTIs in nursing home residents. The Inappropriate Diagnosis of UTI measure determines the proportion of hospitalized medical patients with a positive urine culture who do not meet criteria for UTI and is focused on improving diagnostic accuracy. Data collection burden does not overlap for these measures, as they address different target populations and facilities.

**[Response Ends]**

**5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.**

*Provide analyses when possible.*

**[Response Begins]**
N/A
**[Response Ends]**