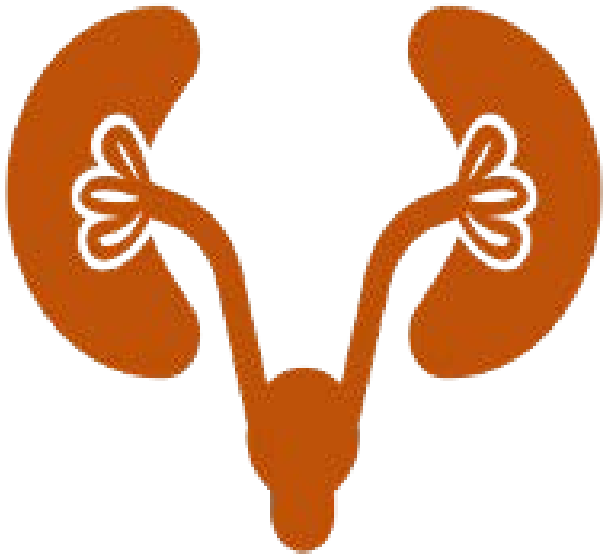
NQF Technical Expert Panel: Summary Report

Inappropriate Diagnosis of Urinary Tract Infection in Hospitalized Patients

December 2021



HMS
MICHIGAN HOSPITAL MEDICINE SAFETY CONSORTIUM

TABLE OF CONTENTS

1 INTRODUCTION

Background

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

1.1 Technical Expert Panel

The TEP was comprised of 11 stakeholders with diverse perspectives and areas of expertise. The panel included experts from a variety of professional organizations, non-profit organizations, and governmental agencies (see table below).

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

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

The first hour of the October 18, 2021, meeting provided a general overview of the Urinary Tract Infection measure that will be submitted to NQF. This was followed by a

presentation and discussion of the data to be included in the submission. A second meeting to review this content was held on October 20, 2021, and contained the same content as the meeting on October 18, 2021. We offered this meeting at two different times to best accommodated attendee schedules and allow for smaller groups to facilitate discussion. The TEP ended with a wrap-up session on October 28, 2021, a two-hour meeting during which we reviewed answer to questions raised during the detailed discussion and elicited final feedback. Each of these sessions were moderated by a physician associated with HMS.

2 DISCUSSION SUMMARY

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

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

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

2.1 Overview of HMS and the Initiative

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

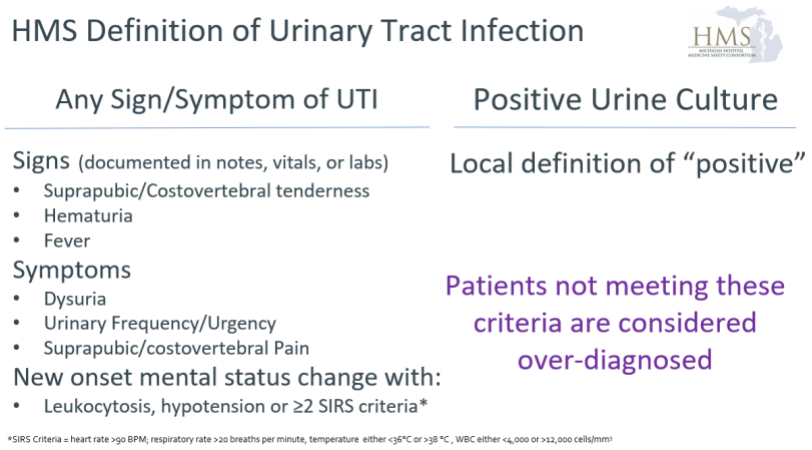
First, the HMS moderator provided rationale for a focus on over-diagnosis. Several case examples were presented in which patients over-diagnosed with an infection they didn't have experienced negative outcomes because of their inappropriate diagnosis, including antibiotic-associated side effects and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

Second, the HMS moderator introduced HMS. HMS is a 62-hospital collaborative quality initiative (CQI) based in the state of Michigan and funded by the Blue Cross and Blue Shield of Michigan. The process for data collection was then outlined. HMS began collecting data on patients hospitalized for diagnosis of UTI in 2017; data were abstracted from medical records in a structured format by trained abstractors, who pseudo-randomly selected up to 8 cases for each measure every two weeks. The selection protocol was also outlined. The HMS dataset currently includes >28,000 patients in the positive urine culture (UTI) cohort. Inclusion and exclusion criteria were reviewed, as was a summary of abstracted data (demographic data, antibiotic information, radiographic findings, microbiologic information, co-morbidities, vitals/oxygen/laboratory information, and symptoms).

Third, the moderator discussed the creation and use of measures of over-diagnosis of UTI and presented conceptual and operationalized definitions. For example, a conceptual definition for over-diagnosis of UTI was devised to identify patients treated for UTI who may not have

UTI. To operationalize the definition, the HMS team relied on national guidelines developed by professional societies.

The operationalized definition was patients with ≥1 sign/symptom of UTI within the 7-day NHSN window period and a positive urine culture, treated with antibiotics for UTI (and without a concomitant infection). The signs and symptoms of a UTI were defined, as were additional exclusion criteria (e.g., delirium without other signs of systemic infection or falls without localizing symptoms).



Data on UTI has been reported to HMS hospitals as quality measures since 2017, and the over-diagnosis of UTI measure has been used as a "pay for performance" metric since 2018. While hospitals are asked to review cases of over-diagnosis at their institutions to drive quality improvement, types of improvement initiatives implemented are at the discretion of the hospital.

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

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

2.2 Introduction of the UTI Measure

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

Detailed UTI content was presented in two sessions, on 10/18/2021 and 10/20/2021. The content presented at both meetings was similar and is reflected in the same summary. The comments of the TEP members from both panels are presented together.

2.2.1 Summary of Presentation

The HMS moderator laid the groundwork for the conversation by identifying the volume of patients currently treated for UTI. As of 2011, there were ~400,000 hospitalizations for UTIs resulting in estimated costs of $2.8 billion. Approximately 1/3 of patients diagnosed with UTI are treated inappropriately with antibiotics. The conversation then focused on the impact of inappropriately treating a patient for a UTI, including identifying which groups are more at risk for being inappropriately diagnosed, and highlighting instances where treatment does not correlate with improved outcomes (patients inappropriately treated with antibiotics have an increased risk of UTI in the following year; they have an increased risk of developing antibiotic resistance and there is a higher risk – especially among the elderly of antibiotic associated adverse-events).

The development of our definitions from a conceptual definition to an operationalized definition was reviewed in more depth. Time was spent reviewing the operationalized definition of UTI. Patients must have at least one sign or symptom of UTI AND have a positive urine culture. Signs and symptoms were defined. Signs included suprapubic/costovertebral tenderness, hematuria, and fever [>38°C]. Symptoms included dysuria, urinary frequency/urgency, and suprapubic/costovertebral pain. We also discussed how we defined patients with altered mental status – to be considered to have a UTI, patients with altered mental status must also have leukocytosis (>10,000 cells/mm^3), hypotension (SBP<90) or ≥2 SIRS criteria. The frequencies of these signs and symptoms were also presented. Positive urine culture was defined based on the hospital's local definition of positive. **Patients who did not meet these criteria were considered to be over-diagnosed with UTI.**

HMS presented the following questions for discussion:

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

2.2.2 TEP Member Discussion

TEP members suggested that the name of the UTI measure needs to be revised, especially if this is to become a national measure. Some of the suggestions included misdiagnosis of UTI, clinical false positive, clinical false discovery rate of UTI resulting in inappropriate treatment, and overtreatment of not UTI; it was also suggested that the name of the measure include treatment or antibiotics. Many members agreed that metrics need to be educational and include phrases that clinicians and patients understand, because it is difficult for a metric to perfectly capture a condition or disease.

Some members in emergency medicine suggested that symptom-based measures like this can be challenging if emergency department (ED) care is driven by consultants. Some members feel this measure would be best to tackle in the ED if possible, noting that cultures are often obtained, and therapy initiated, in the ED. The HMS moderators noted that the measure is developed to not penalize a hospital based on a single dose of antibiotics given in the ED, should those antibiotics be discontinued once the patient is admitted.

The TEP members agreed with using their hospitals' definition of a positive urine culture, with the hope that it would motivate laboratories to focus on greater than a colony forming unit count of $\geq 10^5$. There was also some conversation around how data from a urinalysis was considered, as urologists sometimes utilize the urinalysis to determine presence of infection. There was general agreement, however, that using urine culture was appropriate moving forward.

Some members highlighted the difficulty in treating patients with catheters, as presence or absence of symptoms in this population may be unreliable to diagnose infection. Some suggested it would be helpful to portray the differences more clearly in applicable signs and symptoms in the presence or absence of urinary catheters. In general, TEP members agreed with the exclusion criteria for this measure. One member asked that more detail be added to the altered urinary anatomy population to capture the group more accurately.

There was also conversation around older adults. Many TEP members suggested giving leeway to older patients and avoiding gray area scenarios to remain impactful and give clinicians the benefit of the doubt. Several TEP members provided examples of the literature regarding the relationship between UTI and post-stroke recrudescence and altered mental status in nursing homes. Lastly, some TEP members discussed fever definitions in older adults, noting that there are different potential cutoffs for this population.

2.2.3 Key Takeaways

- The TEP members agreed that the measure title should be tested in workshops and focus groups to lessen any confusion regarding the measure purpose.
    - Response: Potential titles were workshopped and then emailed back to the panel; the eventual final name for the measure was decided to be "inappropriate diagnosis of UTI" (see below for details).
- The TEP members suggested additional analysis be conducted to determine different fever cutoffs for older adult patients (HMS will do additional analysis in this area).

    - Response: Additional data provided to TEP members during follow up session (see details below).
- In general, the TEP members agreed that the current criteria for over-diagnosis of UTI were acceptable.
    - Response: Face Validity Confirmed

2.3 Validity of the UTI Measure

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

2.3.1 Summary of Presentation

After outlining the measure definition, we presented data on validity. First, the HMS moderator discussed the ways HMS assesses patient-level validity. First, HMS provides hospitals with site-specific data reports quarterly and encourages hospitals to discuss cases of "over-diagnosis" with their local experts (e.g., infection control, infectious disease, antibiotic stewardship teams). Hospitals with case clarification concerns are invited to bring cases back to HMS for re-evaluation and updates to measure definitions, if necessary. Due to the long-standing nature of this initiative, however, no recent modifications to the over-diagnosis of UTI measure have been made.

Next, the HMS moderator outlined assessment of patient-level validity through physician implicit review. In 2020, the team conducted formal case reviews that included 2 to 3 physicians, who were asked to review cases sampled from gray areas. Cases within these gray areas were felt to represent those for which classification of over-diagnosis was uncertain and were identified by local hospitals or the HMS data, design, and publications committee. Based on the feedback received from case review, the following changes were included in the final measure:

- symptomatic patients not treated for a UTI will be excluded
- hypogastric pain added as a synonym for suprapubic pain
- patients treated with antibiotics outside the window for collecting symptoms were excluded

The kappa for reviewer agreement was 0.72, indicating substantial agreement.

Next, the team presented data on the measure's content validity, including data on associations between over-diagnosis of UTI and antibiotic overuse, antibiotic associated adverse events, and longer length of stay. We noted that 1/3 of patients were over-diagnosed with a UTI and 83% of these patients were treated with antibiotics for a median of 7 days, with wide variation in treatment of over-diagnosed patients across hospitals. We also noted that treatment of over-diagnosis of UTI was associated with longer duration of hospitalization.

HMS presented the following questions for discussion:

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

2.3.2 TEP Member Discussion

In the conversation regarding whether lower measure scores represent better diagnosis or quality of care, there was a substantial dialogue related to the tactics employed by hospitals to improve their measure scores. The HMS team noted that HMS hospitals used several different initiatives to improve scores, including use of diagnostic stewardship practices, order sets, and prospective audit and feedback. We were unable, however, to attribute success to any single measure. Regarding the types of resources and infrastructure being made available to HMS hospitals, it was noted that nearly all (95%) had a stewardship team. However, it was also recognized that just because a team is in place, does not guarantee a robust stewardship infrastructure. Concerns were raised about how differences in culture and practice patterns would impact outcomes, and about the potential to "game" the system.

One TEP member suggested studying the impact/harms of other conditions or diseases that were "under" diagnosed as a result of being diagnosed incorrectly with UTI. The HMS team noted the data for these analyses were not currently available (nor was the funding to add this to our current model).

There was also conversation relative to the case study analysis the HMS team conducted. In short, there seemed to be agreement about the way these case studies were conducted; it was convincing because the cases that were reviewed were gray areas, where there could have been diagnostic uncertainty.

Finally, there was conversation around urine cultures. While there was agreement that screening based on positive urine cultures was appropriate, screening patients based on urine culture order may allow for inclusion of patients treated for UTI despite negative urine culture. This would expand the patient cohort for which data was available.

2.3.3 Key Takeaways

- There was consensus that the data presented demonstrated strong validity.
    - Response: Face Validity Confirmed
- There was agreement that although using a positive urine culture as the initial screen made sense, we may be missing patients treated inappropriately for UTI who may have had a negative urine culture.

2.4 Reliability of the UTI Measure

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

2.4.1 Summary of Presentation

The HMS moderator introduced the reliability section by discussing how data are collected. All data used in this measure are collected by trained data abstractors, the majority of whom have a nursing background. Each abstractor is provided with an extensive data dictionary outlining established data protocols and charting the data that should be abstracted. These protocols and data include days of hospitalization for which to collect data (i.e., +/- 3 days of positive urine culture), guidance on data entry (range/absolute value), location of data within the medical record (i.e., physician order), definitions for all data elements, and instructions on how to handle missing data.

HMS also has an extensive audit process. Data from each hospital is audited at least once per year. The expectation is that hospitals would attain a ≥95% rate of accuracy. Since the inception of the project, over 300 cases have been audited. The moderator presented data on an additional audit conducted by the HMS program manager. The program manager selected 50 cases chronologically; these cases represented 29 different hospitals. The results of this audit suggest that in 95% of the cases, the presence of a urinary catheter was correctly identified, and that 98.5% of all data was abstracted correctly. When audited data were reapplied to the measure definition, only two of the 50 cases would be reclassified (one changed from over-diagnosis to appropriate diagnosis, one changed from appropriate diagnosis to over-diagnosis). There was an inter-rater reliability of 0.91, which indicates strong to "almost perfect" reliability.

The HMS moderator also reviewed how reliable hospital estimates are at distinguishing variation between hospitals. The HMS moderator presented data on the percent of patients over-diagnosed by hospital and calculated an intraclass correlation of approximately 5%. This allowed us to compute the number of cases that hospitals participating in a measure would need to collect on an annual basis to reach a reliability threshold of 0.8 based on the Spearman Brown Prophecy; n=63. Based on the data currently being collected at HMS, we noted the majority of hospitals would be able to meet an appropriate minimum threshold. All hospitals participating in the collaborative could abstract the necessary number of cases to achieve a reliability of 0.7.

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

We then opened the meeting up for general conversation.

2.4.2 TEP Member Discussion

There was discussion around the whether the presented data represented reliability or validity. Ultimately, it was felt that this data may fit better under validity. Regardless, of location, however, TEP members found the data presented to be compelling and felt that data were strong.

One TEP member suggested adding a target or baseline to acknowledge that there may be issues with misdiagnosing UTI. For example – do you want fewer than 20% of patients treated for a UTI to be over-diagnosed? However, another TEP member noted that if a threshold was built in, then there was the possibility to "game" the measure. Other TEP members disagreed with having a threshold for this measure and noted that the number of cases necessary to achieve appropriate reliability was key. If, for example, reliability could be achieved with abstraction of 30 cases, many hospitals, even critical access hospitals, would be able to meet this threshold.

The conversation around starting with the urine culture order was brought up once again in this section. One TEP member suggested starting with a urine culture order (suspected UTI) and an antibiotic initiation as to not hinge on positive or negative results to capture a slightly larger population.

2.4.3 Key Takeaways

- The group believed it might be necessary to move the data on the audits from the reliability section to the validity section.
    - Response: Pertinent data were moved to the validity section of the NQF submission.
- There was consensus that hospitals would be able to collect a minimum threshold of data needed for an acceptable level of reliability (0.8).
    - Response: Evidence of Feasibility

2.5 Feasibility of the UTI Measure

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

2.5.1 Summary of Presentation

The HMS moderator began a review of the feasibility section by noting that the HMS team evaluated several areas where we could pare down the data collected by HMS to streamline the data collection process. First, we evaluated if it was possible to shorten the NHSN symptom window from day -3 to day +3 to day -1 to +2. Second, we assessed frequency of our exclusion criteria to determine if they could be removed without affecting measure outcomes. Third, we reviewed elements used to define over-diagnosis to evaluate how their removal would affect classification. We noted that in this area, we decided to exclude patients with a spinal cord injury from the sample. Finally, we tested our pared down data collection forms at five hospitals in Utah, which determined that reviewing cases for eligibility took one to three minutes and abstracting data took 15 minutes for easy cases to 30 minutes for harder cases. We compared this to the amount of time it would take to abstract information for other NHSN measures requiring case review. We determined our abstraction timeframe was comparable to other measures. Also, based on the Spearman Brown Prophecy data presented earlier, we determined that a threshold of abstracting 63 cases per year (standard threshold of 0.8) was reasonable, as 90% of the hospitals

participating HMS were able meet this threshold. If the threshold was reduced to 0.7, all HMS hospitals met the metric.

The team also briefly reviewed if it was feasible to collect data electronically instead of via manual abstraction. To facilitate this process, we assessed whether discrete data elements could replace the symptom data currently collected through chart review by looking to order indications. In short, we found that more than half of ASB cases did not have an indication in the urine culture, and another quarter were reflex urine cultures; similar numbers were documented for the UTI cohort. We also found that among those with an indication, the most frequently noted indication was abnormal UA, which did not provide concrete information about the patients' symptoms. We presented data demonstrating that indication data was really only accurate in diagnosing a UTI and was not helpful in measuring inaccurately diagnosed UTIs. As such, we determined that the data for this measure still needed to be collected through manual abstraction.

HMS presented the following question for discussion:

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

2.5.2 TEP Member Discussion

Several members agreed that the amount of time it would take to abstract a case is consistent with what was shown, as long as the list is kept simple. There was some concern that while simple cases most likely would not take long to abstract (<15 minutes), longer cases could take more than 30 minutes to abstract. Some TEP members expressed concern regarding smaller hospitals and having issues operationalizing a process for inclusion screening.

There was also robust conversation around estimates of harm secondary to over-diagnosis of UTI, nationally. Some TEP members suggested that understanding the problem on a national scale would strengthen the measure, as our data is derived exclusively from Michigan.

In addition, we sought feedback regarding a proposal to shorten the data collection window from day -1 to +2, which deviates from the currently used NHSN window. We presented data regarding the amount of information obtained on days -3, -2, and +3, after which TEP members agreed with the narrower window for data collection. It was noted that we need to be explicit with reasoning if we deviate from the NHSN window. One potential rationale is that the NHSN window applies to hospitalized patients and does not typically include days prior to hospitalization, as this data is often unavailable. In contrast, the overdiagnosis of UTI measure would apply mostly to patients presenting with community-acquired UTI; thus, the reduced window collection period has face validity and improves feasibility.

2.5.3 Key Takeaways

- The TEP recommended that HMS determine the implications of over-diagnosis of UTI nationally.

    - Response: National estimates provided in importance to report in NQF application; additional data were provided to TEP during follow-up session (see details below).
- TEP members suggested HMS determine how many of these individuals could be helped if these measures were implemented.
    - Response: Provided evidence on use and usability, suggesting that 37% of over-diagnosis of UTI can be eliminated with pay for performance metrics.
- The members advocated narrowing the NHSN UTI collection window to day -1 to +2 (urine culture day 0) with rationale.
    - Response: Window period changed; improved measure feasibility

2.6 Use of the UTI Measure

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

2.6.1 Summary of Presentation

The final section we presented to the TEP was on measure usability. We started by noting that the over-diagnosis of UTI measure has been in use since 2018 as a pay-for-performance metric at HMS. During this time, we have seen a 37% reduction ($p<0.001$) in over-diagnosis of UTI.

HMS presented the following questions for discussion:

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

2.6.2 TEP Member Discussion

While a several TEP members noted it would be worthwhile to take this on, they note that HMS needs to demonstrate a strong value case for this. The TEP members noted they could see how this measure would be of value to the hospitals that were currently participating in HMS. However, they were concerned about future adoption by other hospitals, who are struggling to care for patients during the COVID pandemic. One TEP member highlighted that decreased length of hospital stay for these patients may provide incentive to get hospital financial officers on board with adopting these measures.

TEP members also raised concerns about the inability to automate the measure at this time and believe abstraction of medical records may be difficult for certain elements. There was also concern about being able to get nurses to abstract this data and discussed if it would be possible to train other individuals to abstract.

Other TEP members voiced strong support of the measure in driving improvement in diagnosis and antibiotic use nationally and suggested it would make a good additional to national measures of antibiotic use (e.g., NHSN AUR metrics).

2.6.3 Key Takeaways

- TEP members agreed this measure could be of value but would benefit from added context about the scope and implications of over-diagnosis nationally.
    - Response: Data on national scope provided in "importance" section of NQF submission. General agreement that measure provided value. Additional data provided during TEP follow-up session (see details below).

3. RESPONSES TO ISSUES RAISED DURING THE TEP

3.1 Responses to Issues Raised During the TEP

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

- Evaluate fever cut-offs for adults
    - In the standard data analysis, HMS defines a fever as ≥38.1°C. Because of the way in which fever data are collected in the HMS UTI cohort (36.1 to 38.0; ≥38.1), we were unable to directly test the impact of a lower definition of fever. However, we did analyze for a 2nd measure that HMS is submitting for NQF endorsement, inappropriate diagnosis of community-acquired pneumonia. In that analysis, we found that using a lower fever threshold would only impact nine patients, and we decided not to change the threshold for a fever. Thus, we are also not making a change for this metric.

- Determine how many patients could be helped if measures were implemented
    - This is not something that we are directly able to do with our data, so we had to turn to the literature. To this end, we looked at number of patients annually that were hospitalized for UTI in the United States and found this to be approximately 400,000 patients. Based on the percent of patients that we know are over-diagnosed from our own work – 1/3 of patients, we can extrapolate that 133,000 hospitalized patients would be over-diagnosed with UTI annually. We suspect this is an underestimate due to under-coding of UTI in diagnostic codes. **These data were added to the NQF submission.**

- Determine implications of over-diagnosis of UTI nationally
    - One implication of over-diagnosis is length of stay. Research suggests that patients over-diagnosed with UTI have an average on 1 day longer length of stay, as well as greater antibiotic associated adverse-events and antibiotic resistance. **These data were added to the NQF submission.**