NATIONAL QUALITY FORUM

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: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 2063

Measure Title: Performing cystoscopy at the time of hysterectomy for pelvic organ prolapse to detect lower urinary tract injury

Measure Steward: American Urogynecologic Society

Brief Description of Measure: Percentage of patients who undergo cystoscopy to evaluate for lower urinary tract injury at the time of hysterectomy for pelvic organ prolapse.

Developer Rationale: Universal intraoperative cystoscopy with vaginal prolapse repair is not consistent across providers performing these procedures. This quality measure will quantify the percentage of women undergoing vaginal prolapse repair who have an intraoperative cystoscopy done to detect intraoperative injury, with the goal of making universal cystoscopy standard for all vaginal prolapse procedures and preventing missed lower urinary tract injury. The measure will quantify the performance gap in pelvic surgeons' use of universal intraoperative cystoscopy and highlight those surgeons' care where a higher level of quality is easily obtained. Patient care will be improved by avoiding unrecognized injury to the lower urinary tract during hysterectomy and vaginal prolapse surgery, thus preventing a highly morbid complication and reoperation for unrecognized injury.

Cost-effectiveness of universal cystoscopy at the time of hysterectomy and vaginal prolapse repair has been demonstrated cost-savings above a threshold incidence of ureteral injury of 1.5%. Ureteral injury at the time of hysterectomy and prolapse repair occurs with incidence up to 5.1%. Universal cystoscopy is cost-effective in preventing unrecognized lower urinary tract injuries. Universal cystoscopy at the time of hysterectomy and vaginal prolapse repair will prevent unnecessary morbidity in unrecognized injury, will prevent hospital readmission, will prevent re-operation, and will save health care dollars.

Numerator Statement: Numerator is the number of patients in whom an intraoperative cystoscopy was performed to evaluate for lower urinary tract injury at the time of hysterectomy for pelvic organ prolapse.

Denominator Statement: The number of patients undergoing hysterectomy for pelvic organ prolapse (identified by CPT codes for hysterectomy and ICD9/10 diagnoses of prolapse as listed in S.9).

Denominator Exclusions: There are no exclusions from the target population.

Measure Type: Process

Data Source: Paper Medical Records, Registry Data

Level of Analysis: Clinician : Group/Practice, Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Nov 12, 2014 **Most Recent Endorsement Date:** Nov 12, 2014

Staff Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is 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 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 of prior review in 2014

Evidence submitted in the prior review year was based on a systematic review of detection of urinary tract injuries during gynecologic surgery which suggest that routine cystoscopy identifies up to five fold more injuries than non-routine cystoscopy.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- Updated evidence included a systematic review and meta-analysis that found that although routine cystoscopy increases identification of urinary tract injuries intraoperatively, there is not a meaningful reduction in urinary tract injuries detected postoperatively. (Evidence not graded)

Questions for the Committee:

- *What is the relationship between performing a cystoscopy to evaluate for lower urinary tract injury at the time of hysterectomy for pelvic organ prolapse and unnecessary morbidity in unrecognized injury, preventing hospital readmissions, and preventing re-operation? How strong is the evidence for this relationship?*

Guidance from the Evidence Algorithm: Process measure based on SR and clinical practice guideline (Box 3) → Specific information on QQC provided (Box 4) → SR not graded, high quantity of evidence but low quality (Box 6) → Low

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

RATIONALE: Systematic review is about measure focus but does not support the measure as the authors concluded that "although routine cystoscopy clearly increases the intraoperative detection rate of urinary tract injury…it does not appear to have much effect on the postoperative injury detection rate".

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer reports that this is the measure's first year in use. Current performance scores from the AQUIRE registry for 503 patients (denominator) and 16 providers who submitted 2017 data to MIPS range from 88.24% to 100% (mean=96.9%, SD=0.057).
  - Overall registry average (includes providers who did not submit data to MIPS): 94.7%

Disparities

- The developer reports there are no disparities between gender since the measure applies only to women. They also note that the average age of the patients in the registry is over 60 and that differences in age would not be statistically significant. The AQUIRE registry does not collect data on race, ethnicity, disability or socioeconomic status.

Questions for the Committee:

- *Does the data provided demonstrate a sufficient gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in performing cystoscopy to evaluate for lower urinary tract injury at the time of hysterectomy for pelvic organ prolapse?*

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

Rationale: Maintenance measure with data demonstrating little opportunity for improvement; no disparities data.

Committee Pre-evaluation Comments: Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing
2b. Validity: Testing; Exclusions**;** Risk-Adjustment**;** Meaningful Differences**;** Comparability**;** Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

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. For maintenance measures – less emphasis if no new testing data provided.

Validity

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. For maintenance measures – less emphasis if no new testing data provided.

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

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

Evaluators: Staff NQF

Evaluation of Reliability and Validity: Staff PA

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *The staff is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *The staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

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

Staff Evaluation of Scientific Acceptability

Measure Number: 2063

Measure Title: Performing cystoscopy at the time of hysterectomy for pelvic organ prolapse to detect lower urinary tract injury

Scientific Acceptability: 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 subcriteria for both reliability and validity to pass this criterion

Instructions for filling out this form:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions. ***Directives that require you to skip questions are marked in red font.***
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- For several questions, we have noted which sections of the submission documents you should ***REFERENCE*** and provided ***TIPS*** to help you answer them.
- ***It is critical that you explain your thinking/rationale if you check boxes that require an explanation.*** Please add your explanation directly below the checkbox in a different font color***.*** Also, feel free to add additional explanation, even if you select a checkbox where an explanation is not requested (if you do so, please type this text directly below the appropriate checkbox).
- ***Please refer to the*** Measure Evaluation Criteria and Guidance document ***(pages 18-24) and the 2-page*** Key Points document ***when evaluating your measures***. This evaluation form is an adaptation of Alogorithms 2 and 3, which provide guidance on rating the Reliability and Validity subcriteria.
- ***Remember*** that testing at either the data element level **OR** the measure score level is accepted for some types of measures, but not all (e.g., instrument-based measures, composite measures), and therefore, the embedded rating instructions may not be appropriate for all measures.
- ***Please base your evaluations solely on the submission materials provided by developers.*** NQF strongly discourages the use of outside articles or other resources, even if they are cited in the submission materials. If you require

further information or clarification to conduct your evaluation, please communicate with NQF staff (methodspanel@qualityforum.org).

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   ***NOTE**: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation. **TIPS**: Consider: Are all the data elements clearly defined? Are all appropriate codes included? Is the logic or calculation algorithm clear? Is it likely this measure can be consistently implemented?*

   ☒Yes (go to Question #2)

   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

   REFERENCE: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

   ***TIPS**: Check the "NO" box below if: only descriptive statistics are provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e., data source, level of analysis, included patients, etc.)*

   ☒Yes (go to Question #3)

   ☐No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified **OR** there is no reliability testing (please explain below, skip Questions #3-8, then go to Question #9)

3. Was reliability testing conducted with computed performance measure scores for each measured entity?

   REFERENCE: "Testing attachment_xxx", section 2a2.1 and 2a2.2

   ***TIPS**: Answer no if: only one overall score for all patients in sample used for testing patient-level data*

   ☐Yes (go to Question #4)

   ☒No (skip Questions #4-5 and go to Question #6)

4. 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.*

   REFERENCE: Testing attachment, section 2a2.2

   ***TIPS**: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.*

   ☐Yes (go to Question #5)

   ☐No (please explain below, then go to question #5 and rate as INSUFFICIENT)

5. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

   REFERENCE: Testing attachment, section 2a2.2

   ***TIPS**: Consider the following: Is the test sample adequate to generalize for widespread implementation? Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

   ☐High (go to Question #6)

   ☐Moderate (go to Question #6)

   ☐Low (please explain below then go to Question #6)

   ☐Insufficient (go to Question #6)

6. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

   REFERENCE: Testing attachment, section 2a2.

***TIPS**: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" go to Question #9)*

☒Yes (go to Question #7)

☐No (if there is score-level testing that you rated something other than INSUFFICIENT in Question #5, skip questions #7-9, then go to Question #10 (OVERALL RELIABILITY); otherwise, skip questions #7-8 and go to Question #9)

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

REFERENCE: Testing attachment, section 2a2.2

***TIPS**: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements*

Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

☒Yes (go to Question #8)

☐No (if no, please explain below, then go to Question #8 and rate as INSUFFICIENT)

8. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?

REFERENCE: Testing attachment, section 2a2

***TIPS**: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?*

☒Moderate (skip Question #9 and go to Question #10 (OVERALL RELIABILITY); if score-level testing was NOT conducted, rate Question #10 as MODERATE)

☐Low (skip Question #9 and go to Question #10 (OVERALL RELIABILITY); if score-level testing was NOT conducted, rate Question #10 (OVERALL RELIABILITY) as LOW)

☐Insufficient (go to Question #9)

9. Was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

REFERENCE: testing attachment section 2b1.

NOTE: Skip this question if empirical reliability testing was conducted and you have rated Question #5 and/or #8 as anything other than INSUFFICIENT)

TIP:** You should answer this question <u>ONLY</u> if score-level or data element reliability testing was NOT conducted or if the methods used were NOT appropriate. For most measures, NQF will accept data element validity testing in lieu of reliability testing—but **check with NQF staff before proceeding, to verify.

☐Yes (go to Question #10 and answer using your rating from <u>data element validity testing</u> – Question #23)

☐No (please explain below, go to Question #10 (OVERALL RELIABILITY) and rate it as INSUFFICIENT. Then go to Question #11.)

OVERALL RELIABILITY RATING

10. **OVERALL RATING OF RELIABILITY** taking into account precision of specifications (see Question #1) and <u>all</u> testing results:

☐High (NOTE: Can be HIGH <u>only if</u> score-level testing has been conducted)

☒Moderate (NOTE: Moderate is the highest eligible rating if score-level testing has <u>not</u> been conducted)

☐Low (please explain below) [NOTE: Should rate <u>LOW</u> if you believe specifications are NOT precise, unambiguous, and complete]

☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the

data element level is not required, but check with NQF staff]

VALIDITY

ASSESSMENT OF THREATS TO VALIDITY

11. Were potential threats to validity that are relevant to the measure empirically assessed ()?

REFERENCE: Testing attachment, section 2b2-2b6

***TIPS**: Threats to validity that should be assessed include: exclusions; need for risk adjustment; ability to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.*

☒Yes (go to Question #12)

☐No (please explain below and then go to Question #12) [NOTE that ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***]

- There were no exclusions and the measure is not risk adjusted. Data for use of cystoscopy were provided for individual surgeons by percentile. Logistic regression was used to identify differences between high volume, intermediate volume, and low volume surgeons. A random audit of 10% of data or 3 patients (whichever was greater) was completed to assess for missing data. Three patients had missing data which were not related to the data collected to calculate the measure.

12. Analysis of potential threats to validity: Any concerns with measure exclusions?

REFERENCE: Testing attachment, section 2b2.

***TIPS**: Consider the following: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?*

☐Yes (please explain below then go to Question #13)

☐No (go to Question #13)

☒Not applicable (i.e., there are no exclusions specified for the measure; go to Question #13)

13. Analysis of potential threats to validity: Risk-adjustment (this applies to all outcome, cost, and resource use measures and "NOT APPLICABLE" is not an option for those measures; the risk-adjustment questions (13a-13c, below) also may apply to other types of measures)

REFERENCE: Testing attachment, section 2b3.

13a. Is a conceptual rationale for social risk factors included? ☐Yes ☐No

13b. Are social risk factors included in risk model? ☐Yes ☐No

13c. Any concerns regarding the risk-adjustment approach?

***TIPS**: Consider the following: **If measure is risk adjusted**: If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are all of the risk adjustment variables present at the start of care (if not, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model? If a measure is NOT risk-adjusted, is a justification for **not risk adjusting** provided (conceptual and/or empirical)? Is there any evidence that contradicts the developer's rationale and analysis for not risk-adjusting?*

☐Yes (please explain below then go to Question #14)

☐No (go to Question #14)

☒Not applicable (e.g., this is a structure or process measure that is not risk-adjusted; go to Question #14)

14. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

    REFERENCE: Testing attachment, section 2b4.

    ☐Yes (please explain below then go to Question #15)

    ☒No (go to Question #15)

15. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?

    REFERENCE: Testing attachment, section 2b5.

    ☐Yes (please explain below then go to Question #16)

    ☐No (go to Question #16)

    ☒Not applicable (go to Question #16)

16. Analysis of potential threats to validity: Any concerns regarding missing data?

    REFERENCE: Testing attachment, section 2b6.

    ☐Yes (please explain below then go to Question #17)

    ☒No (go to Question #17)

- A random audit of 10% of data or 3 patients (whichever was greater) was completed to assess for missing data. Three patients had missing data which were not related to the data collected to calculate the measure.

ASSESSMENT OF MEASURE TESTING

17. Was empirical validity testing conducted using the measure as specified and with appropriate statistical tests?

    REFERENCE: Testing attachment, section 2b1.

    ***TIPS**: Answer no if: only face validity testing was performed; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).*

    ☒Yes (go to Question #18)

    ☐No (please explain below, then skip Questions #18-23 and go to Question #24)

18. Was validity testing conducted with computed performance measure scores for each measured entity?

    REFERENCE: Testing attachment, section 2b1.

    ***TIPS**: Answer no if: one overall score for all patients in sample used for testing patient-level data.*

    ☒Yes (go to Question #19)

    ☐No (please explain below, then skip questions #19-20 and go to Question #21)

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

    REFERENCE: Testing attachment, section 2b1.

    ***TIPS**: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score*

    ☒Yes (go to Question #20)

☐No (please explain below, then go to Question #20 and rate as INSUFFICIENT)

- "To demonstrate the empiric validity of cystoscopy in detecting injury among patients undergoing hysterectomy for prolapse, chi square tests were used to evaluate the differences" between (1) the percentage of patients who have an injury detected compared among those who do not have concurrent cystoscopy; (2) readmission rates due to all causes among those who did and did not have cystoscopy; and (3) rate of readmission among those who do and do not have a lower urinary tract injury detected with intra-operative cystoscopy.

20. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

    ☐High (go to Question #21)

    ☒Moderate (go to Question #21)

    ☐Low (please explain below then go to Question #21)

    ☐Insufficient (go to Question #21)

- Data on 638 patient records. Cystoscopy was performed in 84.5% of procedures. Women who had cystoscopy were more likely than those who did not have cystoscopy to have an injury detected (6.9% of women who had cystoscopy and 0% of those who did not). Readmission rates due to all causes did not differ among women who did and did not have cystoscopy (4.8% vs 5.1%) and the readmission rate among women who had a lower urinary tract injury was lower than that observed among those who did not have an injury (2.7% vs 5%).

21. Was validity testing conducted with <u>patient-level data elements</u>?

    REFERENCE: Testing attachment, section 2b1.

    ***TIPS**: Prior validity studies of the same data elements may be submitted*

    ☐Yes (go to Question #22)

    ☒No (if there is score-level testing that you rated something other than INSUFFICIENT in Question #20, skip questions #22-25, and go to Question #26 (OVERALL VALIDITY); otherwise, skip questions #22-23 and go to Question #24)

22. 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.*

    ***REFERENCE:** Testing attachment, section 2b1.*

    ***TIPS**: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.*

    Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)

    ☐Yes (go to Question #23)

    ☐No (please explain below, then go to Question #23 and rate as INSUFFICIENT)

23. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?

    ☐Moderate (skip Questions #24-25 and go to Question #26)

    ☐Low (please explain below, skip Questions #24-25 and go to Question #26)

    ☐Insufficient (go to Question #24 only if no other empirical validation was conducted OR if the measure has <u>not</u> been previously endorsed; otherwise, skip Questions #24-25 and go to Question #26)

24. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?

    NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary; you should skip this question and Question 25, and answer Question #26 based on your answers to Questions #20 and/or #23]

    REFERENCE: Testing attachment, section 2b1.

    ***TIPS**: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.*

    ☐Yes (go to Question #25)

    ☐No (please explain below, skip question #25, go to Question #26 (OVERALL VALIDITY) and rate as INSUFFICIENT)

25. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

    REFERENCE: Testing attachment, section 2b1.

    ***TIPS**: Face validity is no longer accepted for maintenance measures unless there is justification for why empirical validation is not possible and you agree with that justification.*

    ☐Yes (if a NEW measure, go to Question #26 (OVERALL VALIDITY) and rate as MODERATE)

    ☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, go to Question #26 (OVERALL VALIDITY) and rate as INSUFFICIENT; otherwise, rate Question #26 as MODERATE)

    ☐No (please explain below, go to Question #26 (OVERALL VALIDITY) and rate AS LOW)

OVERALL VALIDITY RATING

26. **OVERALL RATING OF VALIDITY** taking into account the results and scope of all testing and analysis of potential threats.

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

    ☒Moderate (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

    ☐Insufficient (if insufficient, please explain below) [NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT—please check with NQF staff if you have questions.]

Committee Pre-evaluation Comments: Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent
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 developer reports that data are abstracted from a record by someone other than the person obtaining the original information and that some providers enter data into the registry themselves.
- The developer reports that all data elements are in defined fields in electronic claims and that the AQUIRE registry is free to providers/surrogates.
- The developer has made updated the registry to "make the cystoscopy question more obvious (different color) and different placement" for providers who were inadvertently skipping the cystoscopy question.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

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

Committee Pre-evaluation Comments: Criteria 3: Feasibility

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

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

4a. Use evaluate 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**

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

The developer reports that data from the AQUIRE Quality Improvement registry will be available for public reporting on Physician Compare. AQUIRE is approved as a QCDR by the Centers for Medicare & Medicaid Services.

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

- The developer reports that that provides have dashboards showing their performance in real time. Providers may also receive a report and validated score from CMS or call a registry hotline.
- The developer reports that they are soliciting feedback from those being measure. The feedback is not yet available since this is the first year of MIPS reporting.
- Information about the measures were provided to AUGS members. Providers were able to ask questions about measure use or interpretation at the annual AUGS meeting.

Questions for the Committee:

- *How have the performance results be used to further the goal of high-quality, efficient healthcare?*

- *How has the measure been 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 evaluate 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

- The developer reports there are no improvement data to share as this is the measure's first year in use in the registry, however the developer references a 2017 pilot by the Southern California Permanente Medical Group. The pilot found that "performance varied significantly depending on two factors: whether the surgeon was fellowship-trained in female pelvic medicine and reconstructive surgery, and the volume of hysterectomies performed (these factors correlated with each other)".
- The developer notes that all providers reporting on the measure in AQUIRE "are above 81.8% performance score with an average of 96.9%" which could be considered an improvement compared to the average performance score (81.8%) for medium-volume providers in the pilot study.

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).

- Not applicable

Unexpected findings (positive or negative) during implementation

- Not applicable

Potential harms

- Not applicable

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

Committee Pre-evaluation Comments: Criteria 4: Usability and Use

Criterion 5: Related and Competing Measures

Related or competing measures

- No related or competing measures identified.

Harmonization

- Not applicable

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: June 19, 2018

- **No comments submitted.**

- **NQF members have not submitted a support/non-support choice.**

Developer Submission

Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 2063

Corresponding Measures:

De.2. Measure Title: Performing cystoscopy at the time of hysterectomy for pelvic organ prolapse to detect lower urinary tract injury

Co.1.1. Measure Steward: American Urogynecologic Society

De.3. Brief Description of Measure: Percentage of patients who undergo cystoscopy to evaluate for lower urinary tract injury at the time of hysterectomy for pelvic organ prolapse.

1b.1. Developer Rationale: Universal intraoperative cystoscopy with vaginal prolapse repair is not consistent across providers performing these procedures. This quality measure will quantify the percentage of women undergoing vaginal prolapse repair who have an intraoperative cystoscopy done to detect intraoperative injury, with the goal of making universal cystoscopy standard for all vaginal prolapse procedures and preventing missed lower urinary tract injury. The measure will quantify the performance gap in pelvic surgeons' use of universal intraoperative cystoscopy and highlight those surgeons' care where a higher level of quality is easily obtained. Patient care will be improved by avoiding unrecognized injury to the lower urinary tract during hysterectomy and vaginal prolapse surgery, thus preventing a highly morbid complication and reoperation for unrecognized injury.

Cost-effectiveness of universal cystoscopy at the time of hysterectomy and vaginal prolapse repair has been demonstrated cost-savings above a threshold incidence of ureteral injury of 1.5%. Ureteral injury at the time of hysterectomy and prolapse repair occurs with incidence up to 5.1%. Universal cystoscopy is cost-effective in preventing unrecognized lower urinary tract injuries. Universal cystoscopy at the time of hysterectomy and vaginal prolapse repair will prevent unnecessary morbidity in unrecognized injury, will prevent hospital readmission, will prevent re-operation, and will save health care dollars.

S.4. Numerator Statement: Numerator is the number of patients in whom an intraoperative cystoscopy was performed to evaluate for lower urinary tract injury at the time of hysterectomy for pelvic organ prolapse.

S.6. Denominator Statement: The number of patients undergoing hysterectomy for pelvic organ prolapse (identified by CPT codes for hysterectomy and ICD9/10 diagnoses of prolapse as listed in S.9).

S.8. Denominator Exclusions: There are no exclusions from the target population.

De.1. Measure Type: Process

S.17. Data Source: Paper Medical Records, Registry Data

S.20. Level of Analysis: Clinician : Group/Practice, Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Nov 12, 2014 **Most Recent Endorsement Date:** Nov 12, 2014

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?

1. Evidence and Performance Gap – 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.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

nqf_evidence_attachment_7.1_-1-_AUGS_Measure_2063.docx,NQF_evidence_attachment_04.16.18-636594887681974319.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission? Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2063

Measure Title: Performing cystoscopy at the time of hysterectomy for pelvic organ prolapse to detect lower urinary tract injury

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 10/10/2013

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Outcome: 3 Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured intermediate clinical outcome leads to a desired health outcome.

- Process: 5 a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured structure leads to a desired health outcome.
- Efficiency: 6 evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process: Use of cystoscopy in pelvic organ prolapse repair surgery

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or 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.

Vaginal prolapse repair surgery to suspend the vaginal apex (the top of the vagina) or the anterior part of the vaginal wall (cystocele) and remove the uterus (hysterectomy), involve tissues very close to the bladder and the ureters (tubes connecting the kidneys to the bladder). → Injury to these structures (injury rate of up to 5.1%) is often not identified at the time of surgery → Cystoscopy at the time of surgery can identify ureteral and bladder injury that can then be repaired at the time of surgery → Repair at the time of index surgery decreases patient morbidity, and prevents the need for repeat surgery.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of 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. (IOM)

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ S Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☒ Other

Source of Systematic Review: • Title • Author • Date • Citation, including page number URL	• **Urinary tract injury at benign gynecologic surgery and the role of cystoscopy: a systematic review and meta-analysis** • **Teeluckdharry B, Gilmour D, Flowerdew G.** • **December 2015** • **Obstet Gynecol. 2015 Dec;126(6):1161-9** • **https://journals.lww.com/greenjournal/Abstract/2015/12000/Urinary_Tract_Injury_at_Benign_Gynecologic_Surgery.7.aspx**

Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	From Systematic Review: Although of rare occurrence, there is always an inherent risk of damaging the urinary tract in most major gynecologic surgeries. Up to 75% of ureteric injuries are caused by gynecologic surgery and interestingly, most injuries occur during procedures for benign diseases. The timely detection of a urinary tract injury by cystoscopy intraoperatively allows for immediate referral and repair by a urologist or a urogynecologist during the same surgical procedure. There is up to a fivefold increase in the injury detection rates when cystoscopy is used intraoperatively. However, routine intraoperative cystoscopy does not appear to provide any meaningful reduction in the number of postoperatively detected injuries based on the data in this review. The authors acknowledge that postoperatively detected injuries may have been underreported when routine cystoscopy was not used because of undiagnosed injuries or loss to follow-up. The idea of using cystoscopy as a routine screening tool in major gynecologic procedures continues to intrigue surgeons. However, it remains unclear whether its universal use should be advocated in part due to the relatively low number of studies examined in this systematic review, and the potential cost of implementing routine intraoperative cystoscopy after all major gynecologic surgical procedures. From Guideline: “Is it necessary to perform intraoperative cystoscopy during pelvic organ prolapse surgery? Routine intraoperative cystoscopy during POP surgery is recommended when the surgical procedure performed is associated with a significant risk of injury to the bladder or ureter. These procedures include suspension of the vaginal apex to the uterosacral ligaments, sacrocolpopexy, and anterior colporrhaphy and the placement of mesh in the anterior and apical compartments (93, 94). Intraoperative cystoscopy is performed after completion of POP repair while the patient is still under anesthesia and should include a complete survey of the bladder and assessment of efflux of urine from the ureteral orifices. Identified issues such as no flow or reduced flow from the ureter or an injury to the bladder should be addressed intraoperatively. Delay in recognition of a urinary tract injury may lead to increased morbidity.” “Routine intraoperative cystoscopy during POP surgery is recommended when the surgical procedure performed is associated with a significant risk of injury to the bladder or ureter. These procedures include suspension of the vaginal apex to the uterosacral ligaments, sacrocolpopexy, and anterior colporrhaphy and the placement of mesh in the anterior and apical compartments.”
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	No grade was assigned to the evidence

Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	Level C
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	From the Systematic Review: 79 studies with 50 reporting on the frequency of urinary tract injury during benign gynecologic surgery, 21 on the frequency of urinary tract injury during benign gynecologic surgery with intraoperative cystoscopy, and eight studies dealing with complications of robotic surgery for benign indications. For studies that do not involve cystoscopy at all, only studies with more than 500 patients were selected. Studies that used routine intraoperative cystoscopy, because those are fewer in the literature and many involve fewer than 500 patients, were selected irrespective of the number of patients involved. Excluded were letters to the editor, studies involving only selective cystoscopy in higher risk patients, case reports, and reports in which injuries resulting from benign gynecologic surgery could not be distinguished from injuries resulting from obstetric or oncologic procedures From the Guideline: N/A
Estimates of benefit and consistency across studies	The proportion of ureteric and bladder injuries detected intraoperatively without routine cystoscopy is approximately 18% and 79%, respectively. However, when cystoscopy is performed, the proportion of ureteric or bladder injuries detected intraoperatively increases to approximately 95%. Bladder injuries are up to 15 times more likely to be detected intraoperatively when compared with ureteric injuries, although cystoscopy also increases their intraoperative detection rate.
What harms were identified?	The authors specifically note that cystoscopy is not 100% sensitive or specific. Injuries of thermal nature, secondary to devascularization or suture necrosis, can still be missed intraoperatively by cystoscopy, even with visualization of ureteric jets or an intact bladder. Additionally, at least in North America, because of the amount of gynecologic surgery performed yearly, there could be significant costs for a policy of routine intraoperative cystoscopy after all major gynecologic surgical procedures. The authors caution that the cost of routine intraoperative cystoscopy after all major gynecologic surgical procedures could outweigh the benefit if ureteric injury rate does not exceed 1.5% for abdominal hysterectomy and 2% for vaginal or laparoscopically assisted vaginal hysterectomy.

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Chi AM, Curran DS, Morgan DM, et al. Universal cystoscopy after benign hysterectomy: examining the effects of an institutional policy. Obstet Gynecol 2016;127(2):369–375. Acknowledging that cystoscopy is not 100% sensitive, the authors found that institutional implementation of a universal cystoscopy policy at the time of benign hysterectomy was associated with a significant decrease in delayed postoperative urologic complications and an increase in cystoscopic detection of urologic injuries. The authors conclude that it is feasible to train providers at a variety of skill levels to competently perform cystoscopy. This is in contrast to some concerns raised by the previous systematic review. Although the authors note that hospital charges nearly double if a delayed urologic injury occurs, they are unable to definitively answer the question of whether universal cystoscopy at the time of hysterectomy is a cost-effective intervention due to inadequate data regarding not only hospital cost but also psychological and socioeconomic costs of injury. However, the authors do feel that the study cited in the previous systematic review is an underestimate of the cost-effectiveness cutoff since it fails to account for these factors. Milani R, Manodoro S, Cola A, Palmieri S, Frigerio M. Management of unrecognized bladder perforation following suburethral tape procedure Int J Gynaecol Obstet. 2018 Mar 25. ePub ahead of print. No change to conclusions Melon J, Kelly EC, van Delft KWM. Cystourethroscopy following midurethral slings: is it always necessary? Int Urogynecol J. 2018 Mar 21. ePub ahead of print This paper reviews the arguments for and against cystourethroscopy to detect lower urinary tract injury following transobturator and single-incision slings. No conclusions. Cohen SA, Carberry CL, Smilen SW. American Urogynecologic Society Consensus Statement: Cystoscopy at the Time of Prolapse Repair. Female Pelvic Med Reconstr Surg. 2018 Jan 24. ePub ahead of print. No change to conclusions: “Cystoscopy represents a low-risk tool that can be used to decrease the rate of delayed recognition of lower genitourinary tract injuries after pelvic organ prolapse surgery.”

1a.4 OTHER SOURCE OF EVIDENCE

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

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

Universal intraoperative cystoscopy with vaginal prolapse repair is not consistent across providers performing these procedures. This quality measure will quantify the percentage of women undergoing vaginal prolapse repair who have an intraoperative cystoscopy done to detect intraoperative injury, with the goal of making universal cystoscopy standard for all vaginal prolapse procedures and preventing missed lower urinary tract injury. The measure will quantify the performance gap in pelvic surgeons' use of universal intraoperative cystoscopy and highlight those surgeons' care where a higher level of quality is easily obtained. Patient care will be improved by avoiding unrecognized injury to the lower urinary tract during hysterectomy and vaginal prolapse surgery, thus preventing a highly morbid complication and reoperation for unrecognized injury.

Cost-effectiveness of universal cystoscopy at the time of hysterectomy and vaginal prolapse repair has been demonstrated cost-savings above a threshold incidence of ureteral injury of 1.5%. Ureteral injury at the time of hysterectomy and prolapse repair occurs with incidence up to 5.1%. Universal cystoscopy is cost-effective in preventing unrecognized lower urinary tract injuries. Universal cystoscopy at the time of hysterectomy and vaginal prolapse repair will prevent unnecessary morbidity in unrecognized injury, will prevent hospital readmission, will prevent re-operation, and will save health care dollars.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, 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 (4b1) under Usability and Use.*

Updated 06/05/18

There is no performance score for the measures over time as this is its first year of use. Current performance scores from the AQUIRE registry for providers who submitted data to MIPS range from 88.24% to 100% with an average score of 96.9%, standard deviation=0.057. The total number in the denominator is 503 (for providers who submitted data to MIPS) and all data is from 2017, entered into the AQUIRE registry by providers or their surrogates. This measure can only be applied to each patient once as it is part of a hysterectomy, so 503 is not only the number of measured entities but also the number of patients. The overall registry average (including providers who did not submit data to MIPS) is 94.79%. There is no benchmark for this measure as it is new and did not have 20 providers submit at least 20 cases each; we do not have deciles set for the measure.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, 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.

1. Current cystoscopy practice of recent graduates of obstetrics and gynecology residency: a survey study. Nosti PA, Isaacson MA, Iglesia CB. J Reprod Med. 2011 Sep-Oct;56(9-10):373-5.

2. Credentialing residents for intraoperative cystoscopy. Hibbert ML, Salminen ER, Dainty LA, Davis GD, Perez RP. Obstet Gynecol. 2000 Dec;96(6):1014-7

3. Credentialing criteria for ACGME accredited residency programs include cystoscopy as a core skill in urology, obstetrics and gynecology, and fellowship training in urogynecology.

1b.4. 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. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) 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 (4b1) under Usability and Use.*

Updated 06/05/18

The AQUIRE Registry does not collect data on patient race, ethnicity, disability, or socioeconomic status, so we cannot comment on disparities within those subpopulations. The measure only applies to women, so there cannot be differences between genders. The average age of patients in the registry is over 60, so any differences by age would not be statistically significant.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, 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 1b.4

Updated 06/05/18

No data in the literature addresses this specifically. Again, it is only performed on women, mostly mature women, so the opportunity for differences in age and gender are limited/nonexistent. As for race, ethnicity or socioeconomic status, no studies directly address these subpopulations. One study indicates that there is no difference among patient populations of different body mass but otherwise studies have not addressed subpopulations.

2. Reliability and Validity—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.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

Genitourinary (GU)

De.6. Non-Condition Specific*(check all the areas that apply):*

Safety, Safety : Complications

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Elderly

S.1. Measure-specific Web Page *(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.)*

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

No data dictionary **Attachment:**

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

No changes

S.4. Numerator Statement *(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.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Numerator is the number of patients in whom an intraoperative cystoscopy was performed to evaluate for lower urinary tract injury at the time of hysterectomy for pelvic organ prolapse.

S.5. Numerator Details *(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 S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The number of patients undergoing hysterectomy for pelvic organ prolapse (identified by CPT codes for hysterectomy and ICD9/10 diagnoses of prolapse as listed in S.9) who have concomitant cystoscopy identified upon review of the operative report in the electronic medical record or paper chart.

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

The number of patients undergoing hysterectomy for pelvic organ prolapse (identified by CPT codes for hysterectomy and ICD9/10 diagnoses of prolapse as listed in S.9).

S.7. Denominator Details *(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 S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Hysterectomy (identified by CPT codes) performed for the indication of pelvic organ prolapse (identified by supporting ICD9/ICD10 codes)

The prolapse codes for ICD9 -> ICD-10 are, respectively:

618.01 -> N81.11, Cystocele, midline

N81.10, Cystocele, unspecified

618.02 -> N81.12, Cystocele, lateral

618.03 -> N81.0, Urethrocele

618.04 -> N81.6, Rectocele

618.05 -> N81.81, Perineocele

618.2 -> N81.2, Incomplete uterovaginal prolapse

618.3 -> N81.3, Complete uterovaginal prolapse

618.4 -> N81.4, Uterovaginal prolapse, unspecified

618.6 -> N81.5, Vaginal enterocele

618.7 -> N81.89, Old laceration of muscles of pelvic floor

618.81 -> N81.82, incompetence or weakening of pubocervical tissue

618.82 -> N81.83, incompetence or weakening of rectovaginal tissue

618.83 -> N81.84, pelvic muscle wasting

CPT codes for hysterectomy are:

57530 Trachelectomy

58150 Total Abdominal Hysterectomy (Corpus and Cervix), w/ or w/out Removal of Tube(s), w/ or w/out Removal of Ovary(s)

58152 Total Abdominal Hysterectomy (Corpus and Cervix), w/ or w/out Removal of Tube(s), w/ or w/out Removal of Ovary(s), with Colpo-Urethrocystopexy (e.g. Marshall-Marchetti-Krantz, Burch)

58180 Supracervical Abdominal Hysterectomy (Subtotal Hysterectomy), w/ or w/out Removal of Tube(s), w/ or w/out Removal of Ovary(s)

58260 Vaginal Hysterectomy, for Uterus 250 G or Less

58262 Vaginal Hysterectomy, for Uterus 250 G or Less, with Removal of Tube(s), and/or Ovary(s)

58263 Vaginal Hysterectomy, for Uterus 250 G or Less, with Removal of Tube(s), and/or Ovary(s), with Repair of Enterocele

58267 Vaginal Hysterectomy, for Uterus 250 G or Less, with Colpo-Urethrocystopexy (Marshall-Marchetti-Krantz Type, Pereyra Type), w/ or w/out Endoscopic Control

58270 Vaginal Hysterectomy, for Uterus 250 G or Less, with Repair of Enterocele

58275 Vaginal Hysterectomy, with Total or Partial Vaginectomy

58280 Vaginal Hysterectomy, with Total or Partial Vaginectomy, with Repair of Enterocele

58290 Vaginal Hysterectomy, for Uterus Greater than 250 G

58291 Vaginal Hysterectomy, for Uterus Greater than 250 G, with Removal of Tube(s) and/or Ovary(s)

58292 Vaginal Hysterectomy, for Uterus Greater than 250 G, with Removal of Tube(s) and/or Ovary(s), with Repair of Enterocele

58293 Vaginal Hysterectomy, for Uterus Greater than 250 G, with Colpo-Urethrocystopexy (Marshall-Marchetti-Krantz Type, Pereyra Type)

58294 Vaginal Hysterectomy, for Uterus Greater than 250 G, with Repair of Enterocele

58541 Laparoscopy, Surgical, Supracervical Hysterectomy, for Uterus 250 G or Less

58542 Laparoscopy, Surgical, Supracervical Hysterectomy, for Uterus 250 G or Less, with Removal of Tube(s) and/or Ovary(s)

58543 Laparoscopy, Surgical, Supracervical Hysterectomy, for Uterus Greater than 250 G

58544 Laparoscopy, Surgical, Supracervical Hysterectomy, for Uterus Greater than 250 G, with Removal of Tube(s) and/or Ovary(s)

58550 Laparoscopy, Surgical, with Vaginal Hysterectomy, for Uterus 250 G or Less

58552 Laparoscopy, Surgical, with Vaginal Hysterectomy, for Uterus 250 G or Less, with Removal of Tube(s) and/or Ovary(s)

58553 Laparoscopy, Surgical, with Vaginal Hysterectomy, for Uterus Greater than 250 G

58554 Laparoscopy, Surgical, with Vaginal Hysterectomy, for Uterus Greater than 250 G, with Removal of Tube(s) and/or Ovary(s)

58570 Laparoscopy, Surgical, with Total Hysterectomy, for Uterus 250 G or Less

58571 Laparoscopy, Surgical, with Total Hysterectomy, for Uterus 250 G or Less, with Removal of Tube(s) and/or Ovary(s)

58572 Laparoscopy, Surgical, with Total Hysterectomy, for Uterus Greater than 250 G

58573 Laparoscopy, Surgical, with Total Hysterectomy, for Uterus Greater than 250 G, with Removal of Tube(s) and/or Ovary(s)

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

There are no exclusions from the target population.

S.9. Denominator Exclusion Details *(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 S.2b.)*

There are no exclusions from the target population.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including 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 with at S.2b.)*

We do not plan to stratify the results.

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

1. Denominator: Patients of a specific surgeon or group undergoing hysterectomy or trachelectomy for diagnosis of prolapse as defined by CPT and ICD-9/10 codes are identified from administrative data.

2. Numerator: Electronic medical record or paper chart operative notes are reviewed to identify the performance of a cystoscopy at the time of the procedure identified in the denominator.

3. The numerator is divided by the denominator and multiplied by 100 to calculate a percentage (rate/proportion)

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

not applicable

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

not applicable

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Paper Medical Records, Registry Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Clinician : Group/Practice, Clinician : Individual

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Inpatient/Hospital

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

2. Validity – See attached Measure Testing Submission Form

nqf_testing_attachment_7.1_-_AUGS_Measure_2063.docx,nqf_testing_attachment_7.1_4.16.18.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 2063
Measure Title: Performing cystoscopy at the time of hysterectomy for pelvic organ prolapse to detect lower urinary tract injury
Date of Submission: 10/10/2013

Type of Measure:

☐ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

Instructions

1. 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.
2. **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
3. **For outcome and resource use measures**, section **2b3** also must be completed.
4. If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b5** also must be completed.
5. Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
6. If you are unable to check a box, please highlight or shade the box for your response.
7. Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
8. Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
9. 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 version 7.1 of the Measure Testing Attachment.

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.

2a2. Reliability testing [10] 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 [11] 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; [12]

AND

If patient preference (e.g., informed decisionmaking) 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). [13]

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 nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. 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).

11. 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.

12. 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.

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

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

15. 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.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ claims	☐ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

1.2. 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*).

NA-it was not an existing dataset

1.3. What are the dates of the data used in testing? January 1 2007-December 31 2011 and July 2017-December 2017

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.20***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

1.5. How many and which measured entities were 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*)

Since the measure is currently being tracked by the AQUIRE registry, reliability testing was done based on chart review compared to measure calculation in the registry to ensure that 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. Providers who submitted data for MIPS reporting through the AQUIRE Quality Registry were asked to contribute patient charts and operative notes based on their attestation that to CMS that the data was accurate and complete. The institutions

providing data were community health care institutions in the following states: Texas, Alabama, Tennessee, North Carolina, New Mexico, Kansas, Arizona, Pennsylvania, Indiana, and Michigan.

Sixteen out of 18 MIPS reporting practitioners were asked to provide data (two opted for MIPS reporting after data requests were sent out); 13 provided data with three not responding. In total, these 13 surgeons entered 410 patients into the registry who were eligible for this measure and they provided a representative data set that amounted to 12.7% of the eligible patients.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The dataset included deidentified operative notes and billing information from 62 women identified with ICD-10 codes N81.10, N81.11, N81.12, N81.2, N81.3, N81.4, N81.89 and N81.9 (pelvic organ prolapse) and CPT codes 58150, 58152, 58180, 58260, 58262, 58263, 58267, 58275, 58280, 58290, 58291, 58292, 58293, 58294, 58541, 58542, 58543, 58544, 58550, 58552, 58553, 58554, 58570, 58572, 58573 (hysterectomy).

1.7. 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 reported below.

The sample of cases selected for reliability and validity testing were taken from the larger group of patients and the subsample described above.

1.8 What were 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.

N/A-registry does not collect this data

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level 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*)

To determine how reliable the measure was, we first extracted data from the chart surgeons provided. Data was extracted in two ways: operative notes were reviewed for a description that cystoscopy was performed at the time of hysterectomy for pelvic organ prolapse, and billing documents were also reviewed to ensure that providers were in fact performing a qualified procedure and to check for billing for cystoscopy. The number of patients who underwent hysterectomy for pelvic organ prolapse (excluding those for whom a medical reason not to perform cystoscopy was documented in the notes) is the denominator and those that received cystoscopy is the numerator. These numbers were extracted from operative notes from surgeons.

The operative note performance rates were compared with the self-reported performance rates in the registry. The AQUIRE registry determines the numerator and denominator of the measure: the denominator is the number of patients for whom the provider responds Yes to the question, Was surgery for pelvic organ prolapse performed? and also Yes to the question Was the hysterectomy performed for the indication of uterine/ apical prolapse? The numerator is the number of patients for whom the provider subsequently responds Yes to the question Was an intraoperative

cystoscopy performed to evaluate for lower urinary tract injury? This question only appears in the registry if the provider answers affirmatively to the first two questions. If the provider answers No to the cystoscopy question, he or she is given the opportunity to indicate a medical reason why not, which is the denominator exclusion population.

If the measure as it is recorded in the registry is entirely reliable, meaning it measures exactly what it should, there should be no difference in the data extracted from the operative notes and that calculated by the registry. If the branching logic of the registry is flawed or providers are incorrectly interpreting the question, the measure is unreliable and the operative note data would not match what was calculated by the registry.

We therefore calculated the physician-to-physician variance for the data in the registry and the same variance for the data abstracted from charts. If the measure is reliable, the variance should be the same within the registry data and the chart data even if performance differences exist among physicians.

A high reliability number indicates that there is very little difference between data from the registry and data from the sample of charts reviewed therefore the measure in the registry is producing the same results a high proportion of the time when assessed in the same population in the same time period.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Physician to physician variance was highly similar within the registry dataset (sample variance=0.0012222; reliability=1/variance=818.197) and the chart review dataset (sample variance=0)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Reviewing the documentation in the operative note is a highly reliable method to determine if a cystoscopy was performed at the hysterectomy

---

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

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

2b1.2. 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)*

Several studies have demonstrated that routine, as opposed to selective, cystoscopy, will identify more injuries to the lower urinary tract after hysterectomy, a strong indication of this measure's face validity. This is supported by literature as noted in the evidence form.

To demonstrate the empiric validity of cystoscopy in detecting injury among patients undergoing hysterectomy for prolapse, chi square tests were used to evaluate the differences with respect to the following:

1. The percentage of patients who have an injury detected compared among those who do and do not have concurrent cystoscopy.
2. Readmission rates due to all causes among those who did and did not have cystoscopy

3. Rate of readmission among those who do and do not have a lower urinary tract injury detected with intra-operative cystoscopy.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Cystoscopy was performed in 84.5% (539/638). Bladder and/or ureteral injury were identified in 5.8% (37/638).

1. Women who had cystoscopy were more likely than those who did not have cystoscopy to have an injury detected (6.9 % (37/539) v 0% (0/99), p<.007).
2. The readmission rate due to all causes did not differ among women who did and did not undergo cystoscopy (4.8% (26/539) v 5.1% (5/99), p=.923).
3. The readmission rate among women who had a lower urinary tract injury detected intra-peratively was lower than that observed among those who did not have an injury (2.7% (1/37) v 5.0% (30/601).

Separately similar statistics were calculated for the sample of providers from the registry. Cystoscopy was performed in 98.8% (405/410) of the eligible cases for the 13 providers who shared data. Bladder injury was identified in 1.09% (10/917) cases from these providers. However, due to deidentification of the data it is not possible to know the overlap between these populations.

From the charts provided, cystoscopy was performed in 100% of the cases and bladder injury was identified in 0%. However, this sample size was only 62 patients.

2b1.4. What is 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?*)

Consistent with the literature, this analysis provides evidence that routine use of cystoscopy after hysterectomy for prolapse improves detection of lower urinary tract injury. The readmission rate due to all causes did not differ with respect to the performance of cystoscopy. However, the lower rate of readmission among women who had an injury (a high risk group) compared with those did not have an injury (a lower risk group) is suggestive of how intraoperative cystoscopy decreases post operative morbidity and improves outcomes. There is no reason to believe data from the registry will change this interpretation.

2b2. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section* 2b3

2b2.1. Describe the method of testing exclusions and what it tests (*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*)

2b2.2. What were 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*)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., 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*)

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. 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 controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. 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 related to performance gap in 1b)*

1. The performance scores of the individual surgeons at the 5th, 10th, 25th, 50th, 75th and 95th percentiles are provided for use of cystoscopy at the time of hysterectomy for prolapse.

2. Surgeons were evaluated according to the total number of hysterectomies done for prolapse over a four year period of study. Groups of low, intermediate and high volume surgeons were created as follows:
    - high volume: ≥50 cases(>90th percentile)
    - intermediate volume: 11-49 cases (75th-90th percentile)
    - low volume: ≤10 cases(<75th percentile)

    With adjustment for the severity of prolapse, logistic regression was used to identify differences between these groups of surgeons with respect to the performance of cystoscopy at the time of hysterectomy for prolapse.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *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*)

1. The range of performance scores on an individual surgeon basis was as follows:

Number of cases	Percentile Rank of Surgeons	Percentage of hysterectomies with concomitant cystoscopy	95% Confidence Intervals	
638	5	0%	0	0
	10	3.2%	0	4.2%
	25	8.8%	8.2%	11.1%
	50	29.4%	29.4%	32.3%
	75	40%	38.7%	50%
	95	100%	100%	100%

2. High volume surgeons were more likely than the intermediate and low volume surgeons to perform a colpopexy (97% v 78% v 75%, p=.001). When adjusted for prolapse severity the high volume surgeons were more likely to perform a cystoscopy than either the low volume surgeons (OR 6.3, 95%CI 2.4-16.4) or intermediate surgeons to do so (OR 6.4, 95% CI 2.5-16.1).

2b4.3. What is 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? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The measure is able to identify statistically significant and meaningful differences in performance across measured entities with respect to cystoscopy at the time of hysterectomy for prolapse.

1. The confidence intervals of surgeons at each of these percentile ranks do not overlap indicating significant variation in performance.
2. When analyzed by surgeon volume, there is variation in the rate of cystoscopy at the time of hysterectomy for prolapse.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

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 eMeasures). 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.

2b5.1. 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; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to 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 nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

It is possible that of the 16 surgeons asked to provide data, the 3 who did not (nonresponders) performed significantly worse on this measure. However, examination of their scores on the measure from the registry does not fully support this. The average score of the responders was 98.8% and non-responders was 91.8%.

Within the sample of responders, a random audit of 10% of the data in the registry or 3 patients, whichever was greater, was performed to assess for missing data. As noted previously, the providers were selected based on the fact that all participated in MIPS reporting to CMS through the registry, thus they attested to CMS that their data was complete and accurate.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

The random audit of 10% of the patient data in the registry for each of the 13 surgeons who provided data for the analysis yielded 128 patients. Of those, three had missing data, or 2.3%.

The distribution of this missing data was one patient for each of three providers (as opposed to one provider with multiple cases of missing data).

In all cases, the missing data was not related to measures 2063. The provider failed to answer registry questions but NOT the questions used to calculate the numerator and denominator of measure 2063.

2b6.3. What is 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 nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., 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, provide rationale for the selected approach for missing data*)

From the low percentage of missing data and the fact that it was not related to the measure in question, we conclude that the results of this analysis are not biased. We cannot fully rule out that nonresponders chose not to respond because they felt they performed poorly but statistically this does not affect the reliability or validity of the measure.

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.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

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

If other: Some providers enter data into the registry themselves; in which case they are abstracted by the person obtaining original information

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic claims

3b.2. 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 other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. 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.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

Updated 06/05/18

There are two identified issues with data collection, one of which has been addressed, one we are working on. First, providers were inadvertently skipping the question regarding cystoscopy because it only appears after the questions regarding surgery for pelvic organ prolapse and hysterectomy are answered affirmatively. This has been corrected in the registry to make the cystoscopy question more obvious (different color) and different placement. The second potential issue is that providers/surrogates are waiting until late in the calendar year to input patient data rather than doing it in real time. Our reliability testing indicates that data for this measure is adequately captured in operative notes and billing

documents, but the chance of providers/surrogates entering a procedure in error when doing many at a time post-facto is increased. We have increased our messaging reminding providers to enter patients as they perform surgeries and not wait until December. Thankfully, the time and cost of data collection are minimal (no cost, 3-5 minutes of data entry). We evaluated our data for missing data but it is not a significant concern for this measure (see testing attachment).

3c.2. Describe 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)*.

Updated 06/05/18

The AQUIRE registry is free to providers/surrogates to use and there is no cost for the registry to automatically calculate their performance score. Providers who wish to submit data to MIPS pay a fee of $199 for submission. No other costs are incurred by submission of this measure, and since we have moved away from CPT-based data collection, this is no longer an associated cost.

4. Usability and Use

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.

4a. 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.

4.1. Current and Planned Use

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 performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Payment Program Professional Certification or Recognition Program	Quality Improvement (external benchmarking to organizations) AQUIRE QCDR https://aquire.augs.org/Dashboard/login.aspx Quality Improvement (Internal to the specific organization) AQUIRE QCDR https://aquire.augs.org/Dashboard/login.aspx

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

AQUIRE Quality Improvement registry was developed by AUGS and is used by AUGS members across the country and in Canada for their own benchmarking purposes and for MIPS reporting to CMS; this data will be available for public reporting on Physician Compare. AQUIRE is approved as a QCDR by CMS. The measure is collected as part of the registry, which is free for use to all AUGS members. Currently there are 2352 patients in the registry, entered by 40 physicians. Of those 40, 18 completed MIPS reporting to CMS for 2017. Providers can choose to participate as a group or as individuals.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

N/A

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*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.*)

N/A

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

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.

All providers who use AQUIRE, whether or not they choose to report data to CMS through AQUIRE, have a dashboard that shows their performance for the measure in real time, including patients who are eligible, met, and not met. Providers who choose to report to CMS receive a validated score from CMS as well. Additional assistance is available by calling the registry help line or emailing aquire@augs.org

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

When AQUIRE was launched, information about the measures was provided to all AUGS members. Once AQUIRE was approved as a QCDR, the complete measure specs were publicly posted on AUGS website, where they remain available. In addition, AUGS hosted a dedicated booth at our annual meeting dedicated to AQUIRE for providers to ask questions about measure use or interpretation.

Providers using AQUIRE will also receive feedback 4 times as year, as required by CMS to be a QCDR.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Not available yet: we are now soliciting feedback since the first year of MIPS reporting through AQUIRE has been completed both from providers who reported and those who did not to ensure that the questions regarding the measure and the branching logic behind the questions were clear.

4a2.2.2. Summarize the feedback obtained from those being measured.

N/A

4a2.2.3. Summarize the feedback obtained from other users

N/A

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

N/A

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b 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, what are the reasons? 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.

Updated 06/05/18

There are no data on improvement since this is the first year the measure has been in use. As a pilot, in 2017, the Southern California Permanente Medical Group (SCPMG) evaluated women who underwent hysterectomy for pelvic organ prolapse between January 1 and December 31st 2008 to determine whether concomitant cystoscopy was performed. They found the performance varied significantly depending on two factors: whether the surgeon was fellowship-trained in female pelvic medicine and reconstructive surgery, and the volume of hysterectomies performed (these factors correlated with each other). Generalists and low-volume surgeons scored as low as 68.7% and 68.5% respectively on the cystoscopy measure while fellowship-trained and high-volume providers scored as high as 96% and 94% respectively. Most of the providers from whom we have data on the measure as used in the AQUIRE registry do not qualify as high-volume providers but would be considered medium-volume providers. The average performance score for medium-volume providers in the 2017 study was 81.8%. All providers who reported on the measure through the AQUIRE registry are above 81.8% performance score with an average of 96.9%, which could be considered an improvement compared to the pilot study. However, the providers who submitted measure data through the AQUIRE registry are not part of the SCPMG and we cannot account for differences between the populations of providers.

Based on the high performance of the providers in the AQUIRE registry and the data from the 2017 pilot project, we see the major area of improvement with respect to this measure as increasing performance among generalists/providers who were grandfathered in and did not have fellowship training as well as low-volume surgeons. To do this, we are working on extending use of AQUIRE to other related medical specialty societies and sharing the measure with related registries so that GYNs can also track this data.

4b2. Unintended Consequences

The 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).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

not applicable

4b2.2. Please explain any unexpected benefits from implementation of this measure.

5. Comparison to Related or 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.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

No appendix Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American Urogynecologic Society

Co.2 Point of Contact: Colleen, Hughes, Colleen@augs.org, 301-273-0572-

Co.3 Measure Developer if different from Measure Steward: American Urogynecologic Society

Co.4 Point of Contact: Colleen, Hughes, Colleen@augs.org, 301-273-0572-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

AUGS Quality Working Group:

Roger Goldberg, MD

Stewart Wetchler, MD

Daniel M. Morgan, MD

Mark R. Preston, MD

Samantha J. Pulliam, MD

Sage Claydon, MD

Rony Adams, MD

Doug Hale, MD

Lora Plaskon, MD

As members of the AUGS Quality Working Group, all of these members participated in writing the measure, including the specifications, review of the importance, the measure gap, and the evidence.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2013

Ad.3 Month and Year of most recent revision: 02, 2018

Ad.4 What is your frequency for review/update of this measure? Once a year or more frequently as needed

Ad.5 When is the next scheduled review/update for this measure? 01, 2019

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: