NATIONAL QUALITY FORUM
Driving measurable health improvements together

Measure Worksheet

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.
To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3661
Corresponding Measures:
Measure Title: Mismatch Repair (MMR) or Microsatellite Instability (MSI) Biomarker Testing Status in Colorectal Carcinoma, Endometrial, Gastroesophageal, or Small Bowel Carcinoma
Measure Steward: College of American Pathologists
sp.02. Brief Description of Measure: Percentage of surgical pathology reports for primary colorectal, endometrial, gastroesophageal, or small bowel carcinoma, biopsy, or resection, that contain impression or conclusion of or recommendation for testing of mismatch repair (MMR) by immunohistochemistry (biomarkers MLH1, MSH2, MSH6, and PMS2), or microsatellite instability (MSI) by DNA-based testing status, or both
1b.01. Developer Rationale: The results of MMR/MSI testing of a sample are frequently needed to guide treatment decisions, particularly for patients being considered for checkpoint inhibitor therapy. In the absence of MMR/MSI testing, patients may be treated with chemotherapeutic agents they will not benefit from. MMR/MSI testing is also a crucial prognostic marker to determine the presence of Lynch syndrome, an autosomal dominant genetic disorder that is associated with an increased risk for various cancers. Therefore MMR/MSI testing is critical for prognostic as well as treatment reasons. Pathologists are uniquely well positioned at the time of signing out the surgical pathology report to detail the disposition of MMR/MSI testing for that sample. Referring physicians depend on both the pathologists' interpretations of and any recommendations for tests in order to provide quality patient care. If the status is not indicated in each pathology report for the patient, important tests may be missed, or unnecessary repeat testing may be performed delaying treatment and increasing cost. This measure monitors the success of pathologists in effectively communicating this important information for the purpose of care coordination and efficient use of resources.

---

sp.12. Numerator Statement: Surgical pathology reports that contain impression or conclusion of or recommendation for testing of MMR by immunohistochemistry, MSI by DNA-based testing status, or both
sp.14. Denominator Statement: All surgical pathology reports for primary colorectal, endometrial, gastroesophageal, or small bowel carcinoma, biopsy, or resection
sp.16. Denominator Exclusions: 1) Patients with an existing diagnosis of Lynch Syndrome
2) Squamous cell carcinoma of the esophagus

Measure Type: Process
sp.28. Data Source: Other (Electronic Health Data; Laboratory Information Systems (LIS))
sp.07. Level of Analysis: Clinician: Group/Practice

IF Endorsement Maintenance – Original Endorsement Date:
Most Recent Endorsement Date: Not Applicable

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

sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure 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**

Evidence Summary

- This is a process measure using electronic health data and laboratory data at the group/practice clinician level that assesses the percentage of surgical pathology reports for primary colorectal, endometrial, gastroesophageal or small bowel carcinoma, biopsy or resection, that contain impression or conclusion of or recommendation for testing of mismatch repair (MMR) by immunohistochemistry (biomarkers MLH1, MSH2, MSH6, and PMS2), or microsatellite instability (MSI) by deoxyribonucleic acid (DNA)-based testing status, or both.
- The logic model indicates that the testing for these genetic alterations is critical to guide personalized treatment and assess the risk of cancer progression and additional development.
- The developer cited the MMR and MSI Testing for Immune Checkpoint Inhibitor Therapy: Guideline from the College of American Pathologists (CAP) in Collaboration with the Association for Molecular Pathology, American Society for Clinical Oncology, and Fight Colorectal Cancer for their systematic review.
- The developer noted a total of 103 articles of studies were included in the extraction of data and qualitative analysis. The guidelines provided three recommendations:
    - Colorectal Cancer (CRC) patients being considered for immune checkpoint blockade therapy, pathologists should use MMR immunohistochemistry (IHC) and/or MSI by polymerase chain reaction (PCR) for the detection of DNA mismatch repair defects. Although MMR-IHC or MSI by PCR are preferred, pathologists may use a validated MSI by NGS assay for the detection of DNA mismatch repair defects.
    - In gastroesophageal and small bowel cancer patients being considered for immune checkpoint blockade therapy, pathologists should use MMR-IHC and/or MSI by PCR over MSI by next-generation sequencing (NGS) for the detection of DNA mismatch repair defects. Note: This recommendation does not include esophageal squamous cell carcinoma.

    - In endometrial cancer patients being considered for immune checkpoint blockade therapy, pathologists should use MMR-IHC over MSI by PCR or NGS for the detection of DNA mismatch repair defects.
- All three recommendations were Strong Recommendations and evidence was consistent across studies.
- The developer noted that one potential harm was identified: One type of testing (PCR) is slightly more technically challenging with a slightly higher cost than the other type of testing.

Questions for the Committee:

- What is the relationship of this measure to patient outcomes?
- How strong is the evidence for this relationship?

Guidance from the Evidence Algorithm

Process measure is based off a systematic review with grading (Box 3) → QQC is provided (Box 4) →There is moderate certainty that the net benefit is substantial, and benefits outweigh harm (Box 5) →Moderate

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

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

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

- CAP January-December 2020 data:
    - Twelve reporting entities submitted data on CAP 18. The average score was 78.3% with a standard deviation of 20.9 points. Scores ranged from 40.35% to 100%.
    - Eight reporting entities submitted data on CAP 31 although one did not meet the 20-case minimum. The average score was 77.4% with a standard deviation of 16. Scores ranged from 44.9% to 97.8%.
- The developer mentioned several studies that found that 50% or less of patients underwent MMR/MSI testing.

Disparities

- The developer states that race/ethnicity, gender, insurance and/or socioeconomic status, and disability data elements are not readily available in laboratory information systems and are therefore not captured in the Pathologists Quality Registry. It was determined by feasibility testing that these data elements would not be feasible to collect.
- The developer notes that literature demonstrates that provided disparities exist in testing rates based on race, care setting and insurance status.
    - A 2020 study found that testing rates are lower in community care hospitals than in academic hospitals, and White non-Hispanic patients are more likely to get testing than Black non-Hispanic patients. With respect to insurance: patients with no insurance were tested at 38%, private insurance, 47%. Medicare was only 36% and Medicaid 45%.
    - Similar studies from 2019 found lower rates of testing among racial and ethnic minorities as compared to non-Hispanic white patients, with only 6.0% of African American and 3.1% of Hispanic patients having genetic testing done, despite recommendations for universal screening for colorectal cancer.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?

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

Committee Pre-evaluation Comments:

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- Reporting mismatch or microsatellite instability biomarker status in pathology reports impacts cancer treatment decisions, and the efficiency of starting therapies. Without this information, additional testing may be needing leading to treatment delays.
- There is adequate evidence available in the literature to support the use of this measure. This is a process measure. The measure itself is a direct measure related to the desired outcome, that of either performing or recommending MMR or MSI testing on the following pathology specimens: colorectal cancer, gastroesophageal cancer, endometrial cancer, or small bowel cancer. To support the development of the measure, the authors performed a literature review to determine if there was a relationship between measurement and clinical use in the specified population. This measure is based on strong clinical guidelines recommendations from the following professional organizations: CAP, AMP, and ASCO. Knowledge of these results aids the clinician in identifying patients with Lynch syndrome (prognostic) or patients eligible for specific treatment with Checkpoint inhibitors (therapeutic.) Both of these issues have strong clinical implications and are essential for current patient management.
- Data provided indicate that Colorectal Cancer (CRC) patients being considered for immune checkpoint blockade therapy, pathologists should use MMR immunohistochemistry (IHC) and/or MSI by polymerase chain reaction (PCR) for the detection of DNA mismatch repair defects. In gastroesophageal and small bowel cancer patients being considered for immune checkpoint blockade therapy, pathologists should use MMR-IHC and/or MSI by PCR over MSI by next generation sequencing (NGS) for the detection of DNA mismatch repair defects. In endometrial cancer patients being considered for immune checkpoint blockade therapy, pathologists should use MMR-IHC over MSI by PCR or NGS for the detection of DNA mismatch repair defects. All data was consistent and provided a strong recommendation supporting the measure focus.
- Evidence supports improving the use of the additional tests if certain carcinomas are found by pathologists. It is especially true as different findings suggest different follow-up genetic testing for optimal evaluation.
- MSI/MMR information from tumor samples is needed to make treatment decisions providing the link from process to outcomes; because it's viewed as essential, it wouldn't be viewed as ethical to randomize; studies of the treatments are predicated on the testing being done first
- The evidence appears to favor this measure, although this area is far afield from my expertise--how did it end up with this committee??

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- The range of pathology reports with this information included is 40% to 100% (sd 20.9). It is difficult to uncover race/ethnicity, gender, insurance, or socioeconomic information from laboratory information systems.

- The developers presented evidence of a performance gap. Review of the relevant literature demonstrated a performance gap with very low (often less than 50%) measurement of MMR and/or MSI. They also evaluated data that was available in the Pathology Quality Database for both colorectal cancer specimen testing and endometrial specimen testing. For calendar year 2020, they identified an overall performance in colorectal cancer of 78.3% and endometrial cancer of 77.4% among 12 practices which included a total of 123 pathologists. For the period of 1/21-10/21, an average score of 86.5% of the combined group of colorectal and endometrial cancer was seen. This data suggests a gap in performance but improvement in performance for the groups analyzed. Since this is a voluntary reporting system, it could be that the participants had a higher level of performance. In addition, the measure includes both actual performance or recommendation of doing the testing which suggests that the actual performance of completion of the testing on specimens could be low. There is no evidence provided regarding disparities because this data is not readily available in the PQD and no feasible to collect. There is some suggestion in the literature cited that there are disparities in performance.
- On CAP 18 scores ranged from 40.35-100%, On CAP 31 scores ranged from 44.9 to 97.8%. Literature suggests (6% Black Americans and 3.1% Hispanic Americans received genetic testing) that there are racial, ethnic, and socio-economic status gaps. Data was not available to analyze the study participants by race/ethnicity.
- The gap appears to be in certain sites and certain types of ethnic groups. These data suggest disparities in care that this measure could address.
- Yes, a gap exists: Twelve reporting entities submitted data on CAP 18. The average score was 78.3% with a standard deviation of 20.9 points. Scores ranged from 40.35% to 100%. ○ Eight reporting entities submitted data on CAP 31 although one did not meet the 20-case minimum. The average score was 77.4% with a standard deviation of 16. Scores ranged from 44.9% to 97.8%. • The developer mentioned several studies that found that 50% or less of patients underwent MMR/MSI testing.
- While there is a gap, I am unconvinced this measure warrants a national performance measure.

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.

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.

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: NQF Staff

NQF Staff Evaluation Summary:

Reliability

- Reliability testing at the Accountable Entity level:
  - The developer conducted reliability testing using clinical data from 51 clinicians in academic and private practices of 1282 cases of colorectal, endometrial, gastroesophageal, and small bowel cancer. The 51 volunteer data abstractors were instructed to pull all cases for those diagnoses (per ICD-10 codes) for calendar 2019. No further sample data was extracted.
  - Developers conducted a signal to noise analysis from Rand (2009) using a beta binomial model with the beta distribution defined by alpha and beta to calculate within and between provider variances, which is appropriate for this type of measure. The signal is the proportion of the performance variability that explains actual differences in performance and the noise is the total variability of measured performance.
  - The developer reports a single reliability level (rather than the two specified in the submission). The developer submitted a comment stating that the group score is simply the sum of the individual scores and thus a separate method of calculating a group score was not needed.
  - The mean reliability score was 0.96, with a standard deviation of 0.07, tenth percentile 0.91, ninetieth percentile 1.00. A mean above .7 generally indicates a measure is reliable.

Validity

- Validity testing at the Accountable Entity level:
  - The developers conducted face validity of the computed measure score using 40 subject matter experts (SMEs) including pathologists, gastroenterologists, and genetic counselors.
  - Using and anonymous online survey, SMEs were provided the measure specifications and were asked a single question: "The MMR/MSI Testing Status quality measure as described above will accurately distinguish between good and poor quality of care" on a scale of 1 to 4 where 1 is Strongly Disagree and 4 is Strongly Agree.
  - For this measure, 22 experts answered 4 (Strongly Agree), 15 answered 3 (Agree), 1 answered 2 (Disagree) and 2 answered 1 (Strongly Disagree).
  - The average score for the measure is 3.4, which the developer interprets as the clinicians believing the measure as written will differentiate between good and poor care.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

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

Preliminary rating for reliability: ☐ **High** ☒ **Moderate (Individual Level)** ☐ **Low** ☒ **Insufficient (Clinician: Group/Practice)**

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

Committee Pre-evaluation Comments:

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other

specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- A single reliability level was reported with a mean reliability score of 0.96. It is uncertain if the score is at the individual or clinic group level. Both are required to be reported.
- The data elements are clearly defined. The codes are provided and appear complete. The logic and algorithm for capturing data is clear. I have no concerns that the measure could be reliably implemented. The data is available in multiple electronic sources within the EHR.
- No concerns. Specifications are clear.
- Unclear if Reliability standards have or could be met by the developer of the new measure. Possible discussion of the Committee with the developer may assist in better evaluating the reliability of the measure.
- No concerns
- Appears reasonable.

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- Methodology for reliability testing is appropriate but testing at both clinician and clinical group levels are not presented.
- The NQF reviewers point out that the reliability testing is not adequate because the measure application specifies that the measure is performed at two levels, group and individual, and the reliability testing is provided for only one level which is not clearly specified. It appears that the reliability testing is performed at the individual level and the reliability is very high (.96 for 51 clinicians.) All other descriptions of the data and the methods suggest measurement at the individual level, and it is best to ask the developers if they intended to submit at both levels or at just one level before determining that the measure has insufficient information upon which to make a determination. The methodology was appropriate (ratio-to-noise testing using the Rand methods) and resulted in a high level of reliability and should be able to determine real differences in provider performance.
- The developers report a single reliability level (rather than the two specified in the submission) and do not designate whether it is the individual or clinical group level. ○ The mean reliability score was 0.96, with a standard deviation of 0.07, tenth percentile 0.91, ninetieth percentile 1.0. The measure should be reliable.
- Seems move from group to individual clinician which is confusing. Also appears to have only one measure rather than two or more.
- No concerns
- The lack of reliability testing at the clinician level is a problem.

2b1. Validity -Testing: Do you have any concerns with the testing results?

- No concerns. Face validity is reported.
- No. The measure developers submitted data derived from performing Face Validity testing of 40 experts revealing a high level of Face Validity. They also presented data suggesting performance gaps from the PQD further supporting the validity of this measure. I have no concerns with validity.
- Small number of professionals (40). A small percentage did not respond that the results would distinguish good rom bad care.
- Some concern about the need for manual abstraction but moderate validity testing.
- No concerns
- No, appears to have validity testing.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- Exclusions appropriate. Risk adjustment not done as part of this measure.
- Regarding exclusions: The exclusions are well-defined and appear to be appropriate. I would question whether the exclusion related to whether or not the patient is a candidate for checkpoint inhibitors is a reasonable exclusion since that is not likely to be information that is readily available at the time initial pathology. Please ask the developers for further clarification of the rationale since this is based on clinical conditions that are not generally available to the laboratory. There is no risk adjustment for this measure.
- Exclusions include: 1) Patients with an existing diagnosis of Lynch Syndrome 2) Squamous cell carcinoma of the esophagus. No concerns with exclusions. No risk adjustment for this measure.
- The exclusions need to be explicit as part of the measure. It also helps explain the data and improve reliability.
- Risk adjustment not needed.
- No real evaluation of disparities/risk adjustment

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- Uncertain
- This measure supports the identification of meaningful differences in performance. There was one example suggesting the missing data was low and not likely to affect the validity of this measure evaluating one pathologist's data in the database. There do not appear to be significant threats to validity.
- There were no threats to validity. Data source consistent, data available, no missing data indicated.
- More analysis of these data would be helpful.
- No concerns. If data are missing, then it's not available to clinicians.
- No serious issues.

Criterion 3. Feasibility

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

- Data elements are generated or collected by and used by healthcare personnel during the provision of care.
- All data elements are in defined fields in a combination of electronic sources

- The developer noted that no feasibility issues were identified. The measure is in use in 2021 in the Pathologists Quality Registry. Feasibility testing using a score card was also conducted with pathologists, genetic counselors, and gastroenterologists.

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?
- Is the data collection strategy ready to be put into operational use?

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

Committee Pre-evaluation Comments:

Criteria 3: Feasibility

3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?

- No issues identified, and measure elements are reported in Pathologists Quality Registry.
- This measure appears to be feasible since the data is routinely generated in the course of medical care and is available in an electronic format within the EHR or other electronic sources. I have no concerns with the data collection strategy.
- Data is available in EHR. No concerns with data availability or consistency in data as it comes from standardized fields.
- Appears to be moderate as the Pathology Quality Registry also uses for current MIPS.
- Should be generated during routine care delivery
- Feasible.

Criterion 4: Usability and Use

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**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

Accountability program details

- This measure is currently in use in the Merit-Based Incentive Payment System (MIPS).

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

- This is a new measure, so feedback has not been collected. Feedback will be solicited from those who submit the measure to CMS in early 2022 as part of the regular registry assessment.

Additional Feedback:

N/A

***Questions for the Committee**:*

- How have (or can) 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 evaluates the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

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

Improvement results

- The developer notes that this is a new measure so information about performance improvement or care improvement is not available.

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

Unexpected findings (positive or negative) during implementation

- The developer notes that there were no unexpected findings.

Potential harms Additional Feedback:

- No unintended negative consequences to individuals or populations

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

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

Committee Pre-evaluation Comments:

Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- Has not been accomplished (New measure)
- It is not clear that this measure is being publicly reported. It is being used for MIPs reporting beginning in 2021 and the results of this reporting were not available at the time of the application.
- The measure is not currently in publicly reported programs. The measure is in use in 2021 in the Pathologists Quality Registry. Results not provided.
- A form of this is already in use for pathologists and this would extend the measure to the ordering provider (oncologist, internal medicine, etc.).
- New measure
- Used in pathology QI

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- No reported concerns with usability
- This measure appears to be important for performance improvement since the data derived from MMR and/or MSI reporting are important for both prognostic and therapeutic decisions and would be used to change prevention and therapeutic strategies. Providing clinicians with this information will enhance high quality care and assure patient access to state-of-the-art medical care. There are no apparent harms or unintended consequences. The appropriate use of checkpoint inhibitors can be associated with improved clinical outcomes and inappropriate use can be associated with unacceptable toxicities and no clinical benefits.
- New measure. No unintended consequences noted. Feedback is given to practitioners which may be useful in improving quality.
- Supports high quality healthcare and benefits outweighs harms.
- New measure
- Doubt there are many non-economic harms.

Criterion 5: Related and Competing Measures

Related or competing measures

- No related or competing measures were identified.

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

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- No competing measures
- None identified.
- No related or competing measures.
- None
- N/a
- No

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/12/22

- **Of the zero NQF members who has submitted a support/non-support choice:**
  - **Zero supports the measure**
  - **Zero do not support the measure**
- Comment by: **College of American Pathologists**

For reliability testing, the CAP only performed testing at the individual level. This was for two reasons: first, since the testing we did (signal to noise with a beta-binomial model) is dependent on the number of measured entities, we started with the testing that would yield the lower reliability value, which is testing at the individual level. Given that our individual-level reliability was very high, we did not proceed to group level testing. Second, for purposes of MIPS reporting (which is the only program this measure is for), the group score is simply the sum of the individual scores, there is no separate method of calculating a group score. So calculating "group" reliability doesn't have an independent meaning.

Staff Scientific Acceptability Review

RELIABILITY: SPECIFICATIONS

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

   Submission document: "Measure_3661_MIF" document, items S.1-S.22

2. **Briefly summarize any concerns about the measure specifications.**

RELIABILITY: TESTING

Submission document: "Measure_3661_MIF" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level ☒ Accountable Entity (previous measure score) ☐ Patient/Encounter (previous data element) ☐ Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure ☐ ☒ Yes ☒ No**
   - The developer conducted one level of reliability testing, although the measure is specified for both the individual and clinical group levels of analysis. It is not apparent which level was tested.
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ Yes ☐ No ☒ Not applicable

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

   Submission document: Testing attachment, section 2a2.2

   - Developers conducted a signal to noise analysis from Rand (2009) using a beta binomial model with the beta distribution defined by alpha and beta to calculate within and between provider variances, which is appropriate for this type of measure. The signal is the proportion of the performance variability that explains actual differences in provider performance and the noise is the total variability of measured performance.
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   - The mean reliability score was 0.96, with a standard deviation of 0.07, tenth percentile 0.91, ninetieth percentile 1.00. Assumed estimates based upon beta-binomial model distribution (using methods described by Adams in the 2009 Rand tutorial), alpha=0.2, beta=0.12. The reliability score for this measure is 0.96, very high, which attributes nearly all variability to real differences in performance.

- The raw performance scores had a mean of 60% with a standard deviation of 39 points, tenth percentile 0%, ninetieth percentile 100%, which speak to the potential for improvement in the measure and practice.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

   ☒ **Not applicable** (data element testing was not performed)

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

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

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

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

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

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

- The specifications are precise, unambiguous, and consistently implementable (Box 1) >>> Empirical reliability testing was conducted at the accountable entity level using statistical tests as specified (Box 2) >>> Computed accountable entity testing was not conducted for both individual and clinical groups levels of analysis (Box 4) >>> Reliability testing was not conducted at the patient/encounter level (Box 8) >>> Insufficient

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

- In sp.22, - 4., the developer lists four instances where a case would be removed from the denominator, although these are not list these as exclusions. They include: 1) Insufficient tissue for testing or tissue is necrotic; 2) Specimen contains metastatic carcinoma (not a primary neoplasm); 3) No residual tumor (post treatment); and 4) Patient is not a candidate for checkpoint inhibitor therapy.
- The developer states the exclusions were assessed in the SME analysis, although not specifically identified by question to the SME group. No statistical analysis of exclusions is provided.

Submission document: Testing attachment, section 2b2.

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

Submission document: Testing attachment, section 2b4.

- Simple rates and ranges were provided without conducting statistical testing.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.

- The measure does not have multiple specifications or sources for the data needed to calculate performance.

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

Submission document: Testing attachment, section 2b6.

- The developer states the frequency of the missing data is presumed very low.

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable, the measure is not risk adjusted or stratified

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No ☒ Not applicable

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No ☒ Not applicable

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No ☒ Not applicable

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No ☒ Not applicable

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No ☒ Not applicable

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No ☒ Not applicable

16d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No ☒ Not applicable

16e**. Assess the risk-adjustment approach**

For cost/resource use measures ONLY:

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

☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)** ☒ Not applicable. This is not a cost/resource use measure

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):** NA. This is not a cost/resource use measure

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**

20. **Method of establishing validity of the measure score:**

☒ **Face validity**
☐ **Empirical validity testing of the measure score**
☐ **N/A (score-level testing not conducted)**

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

Submission document: Testing attachment, section 2b2.2

- The developers conducted face validity of the computed measure score using 40 subject matter experts (SMEs) including pathologists, gastroenterologists, and genetic counselors. There was not overlap between the two testing populations of SMEs and clinicians or groups used for reliability testing. Scant details for the SMEs are provided beyond medical specialties and roles.
- Using and anonymous online survey, SMEs were provided the measure specifications and were asked a single question: "The MMR/MSI Testing Status quality measure as described above will accurately distinguish between good and poor quality of care" on a scale of 1 to 4 where 1 is Strongly Disagree and 4 is Strongly Agree. They were also offered the opportunity to provide comments and feedback.
- For this measure, 22 experts answered 4 (Strongly Agree), 15 answered 3 (Agree), 1 answered 2 (Disagree) and 2 answered 1 (Strongly Disagree).

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

Submission document: Testing attachment, section 2b2.3

- The average score for the measure is 3.4, which the developer interprets as the clinicians believing the measure as written will differentiate between good and poor care.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

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

Submission document*: Testing attachment, section 2b1.*

☐ **Yes**

☐ **No**

☒ **Not applicable** (data element testing was not performed)

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

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

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

- Threats to validity were scantly assessed (i.e., exclusions and missing data/non-responders) for this initial measure (Box 1) >>> Empirical validity testing was not conducted (Box 2) >>> Face validity was systematically assessed for the computed measure score by recognized SMEs (Box 3) >>> SME face

validity results indicate substantial agreement the measure score as specified can distinguish quality (Box 4) >>> Moderate

Developer Submission

1. Importance to Measure and Report

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

---

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

2021 Submission:
Updated evidence information here.

2018 Submission:
Evidence from the previous submission here.

Evidence

1a.01. Provide a logic model.

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

[Response Begins]
In 2017, the FDA approved pembrolizumab, a drug to treat cancers with certain genetic markers, irrespective of the cancer site. This represents the first FDA approval of a cancer drug that was not specific to a certain organ or part of the body. Instead, this treatment targets cancers with specific genetic markers and is especially important for patients who have progressed following prior treatment and who have no other alternatives. However, genetic testing of tumors is needed to identify patients for whom the treatment might be effective. This measure quantifies the rate of documentation of the specific genetic alterations that make a patient more or less likely to respond to this cancer drug (documentation includes genetic tests performed, recommended, previously performed, etc.). Since the drug is approved for treatment of multiple types of cancer at multiple anatomic locations, this measure looks at several different forms of cancer. These same genetic alterations have also been shown to be present in certain syndromes that are characterized by an increased risk of cancer in multiple locations at a younger than average age. Therefore, testing for these genetic alterations is critical to guide personalized treatment and assess the risk of cancer progression and additional development.
[Response Ends]

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

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

[Response Begins]

Clinical Practice Guideline recommendation (with evidence review)

[Response Ends]

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

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

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

[Response Begins]

MMR and MSI Testing for Immune Checkpoint Inhibitor Therapy: Guideline from the College of American Pathologists (CAP) in Collaboration with the Association for Molecular Pathology (AMP), American Society for Clinical Oncology (ASCO), and Fight Colorectal Cancer.

This guideline is in submission so no URL is available. If requested, an embargoed copy may be made available.

[Response Ends]

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

[Response Begins]

1. In CRC patients being considered for immune checkpoint blockade therapy, pathologists should use MMR-IHC and/or MSI by PCR for the detection of DNA mismatch repair defects. Although MMR-IHC or MSI by PCR are preferred, pathologists may use a validated MSI by NGS assay for the detection of DNA mismatch repair defects.
2. In gastroesophageal and small bowel cancer patients being considered for immune checkpoint blockade therapy, pathologists should use MMR-IHC and/or MSI by PCR over MSI by NGS for the detection of DNA mismatch repair defects. *Note:* This recommendation does not include esophageal squamous cell carcinoma
3. In endometrial cancer patients being considered for immune checkpoint blockade therapy, pathologists should use MMR-IHC over MSI by PCR or NGS for the detection of DNA mismatch repair defects

[Response Ends]

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

[Response Begins]

Strength of evidence ranges from moderate to low. Moderate certainty: There is moderate confidence that available evidence reflects true effect. Further research is likely to have an important impact on the confidence in estimate of effect and may change the estimate.

Low certainty: There is limited confidence in the estimate of effect. The true effect may be substantially different from the estimate of the effect.

[Response Ends]

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

[Response Begins]
Other grades in the evidence grading system include High and Very Low certainty. High certainty: There is high confidence that available evidence reflects true effect. Further research is very unlikely to change the confidence in the estimate of effect

Very low certainty: There is very little confidence in the estimate of effect. The true effect is likely to be substantially different from the estimate of effect. Any estimate of effect is very uncertain.
[Response Ends]

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

[Response Begins]
All three recommendations are Strong Recommendations: Convincing or adequate strength of evidence; Evidence-to-decision framework values are to the far right or far left (recommending for or against an intervention).
[Response Ends]

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

[Response Begins]
This Guideline uses the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) Framework. Other grades in the recommendation grading system include Conditional Recommendation and Good Practice Statements. Conditional Recommendation: Adequate, inadequate, insufficient strength of evidence; Evidence-to-decision framework values are to the inner right or left (neither recommending for or against an intervention)

Good Practice Statements: Recommendations that panels may consider important but are not appropriate to be formally rated for certainty of evidence
[Response Ends]

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

[Response Begins]
A total of 6642 studies met the eligibility requirements for screening. Based on review of these titles and abstracts, 427 articles met the inclusion criteria and continued to full-text review. From these, 103 articles were included for data extraction and qualitative analysis.

Various types of studies were considered. Higher tiered evidence (clinical practice guidelines, systematic reviews, metanalyses, and randomized control trials) are often not available in the pathology literature. For this Guideline, studies were selected for inclusion if they met the following criteria: 1) the study included human patients; 2) the study population consisted of adult or pediatric patients with advanced solid malignancies being considered for checkpoint inhibitor therapy and adult and pediatric patients with possible Lynch syndrome; 3) the study was published in English; 4) the study compared, prospectively or retrospectively, laboratory testing methodologies for MMR and MSI; 5) the study addressed one of the key questions; 6) the study included measurable data such as diagnostic test characteristics, accuracy of MMR defect detection, survival outcomes or treatment response, germline testing or genetic counseling; 7) studies must be peer-reviewed full-text articles. Conversely, articles were excluded if they were meeting abstracts, noncomparative or qualitative studies (editorials, commentaries, letters, etc.), animal studies only, or articles for which the full text was not available in English. Similarly, studies which addressed the concept of MMR/MSI testing generally but did not address at least one of the key questions with outcomes of interest were omitted.
[Response Ends]

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

[Response Begins]

All three recommendations were Strong Recommendations and evidence was consistent across studies. The Expert Panel for this Guideline concluded for all three recommendations that tests were very accurate and carried large benefits and only small harms. The guidance statements are expected to be acceptable to key stakeholders and feasible to implement.
[Response Ends]

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

[Response Begins]
Only small potential harms were identified. The recommended tests are very accurate, both sensitive and specific. One type of testing (PCR) is slightly more technically challenging with a slightly higher cost than the other type of testing, IHC. A newer type of testing, NGS, can have several advantages over traditional PCR but carries a higher cost and is not available in all laboratories, therefore it was not recommended as a primary method.
[Response Ends]

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

[Response Begins]
N/A; Guideline is not yet published
[Response Ends]

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

Performance Gap

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

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

[Response Begins]

The results of MMR/MSI testing of a sample are frequently needed to guide treatment decisions, particularly for patients being considered for checkpoint inhibitor therapy. In the absence of MMR/MSI testing, patients may be treated with chemotherapeutic agents they will not benefit from. MMR/MSI testing is also a crucial prognostic marker to determine the presence of Lynch syndrome, an autosomal dominant genetic disorder that is associated with an increased risk for various cancers. Therefore MMR/MSI testing is critical for prognostic as well as treatment reasons.

Pathologists are uniquely well positioned at the time of signing out the surgical pathology report to detail the disposition of MMR/MSI testing for that sample. Referring physicians depend on both the pathologists' interpretations of and any recommendations for tests in order to provide quality patient care. If the status is not indicated in each pathology report for the patient, important tests may be missed or unnecessary repeat testing may be performed delaying treatment and increasing cost. This measure monitors the success of pathologists in effectively communicating this important information for the purpose of care coordination and efficient use of resources.
[Response Ends]

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

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

[Response Begins]
For January-December 2020, CAP 33 existed as 2 separate measures (additional diagnoses were added when the two were combined): CAP 18 (Colorectal Cancer) and 31 (Endometrial Cancer). Twelve reporting entities submitted data on CAP 18. The average score was 78.3% with a standard deviation of 20.9 points. Scores ranged from 40.35% to 100%.

Eight reporting entities submitted data on CAP 31 although one did not meet the 20-case minimum. The average score was 77.4% with a standard deviation of 16. Scores ranged from 44.9% to 97.8%.

For January-October 2021, 16 practices have entered data (4 do not meet the 20-case minimum yet). The average score so far is 86.5%. ***NOTE***: this information has been updated since Intent to Submit was sent.
[Response Ends]

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

[Response Begins]
In a 2018 study of 152,993 adult colorectal cancer patients including 17,218 "younger" adult patients, only 28.2% and 43.1% respectively underwent MMR deficiency testing (1). In a 2019 study of clinicians, 249 participants including 237 gastroenterologists responded to a survey. Less than one-third of practicing gastroenterologists indicated that their colorectal cancer patients were undergoing routine MMR/MSI testing. Specifically of US-based clinicians in this study, 42% (65/153) indicated universal MMR/MSI testing for colorectal cancer patients (2).

In a 2018 survey of clinicians, 50% of survey respondents did reflex genetic testing for endometrial carcinoma (3).

A 2017 study found that MMR/MSI testing was performed in only 51% of studies of gastric carcinoma. The number of studies that included results from all four MMR proteins (as is recommended) was considerably smaller, at only 14% (4). A 2017 study that was a secondary analysis of the MAGIC (Medical Research Council Adjuvant Gastric Infusional Chemotherapy) trial found that of the 503 participants, any genetic testing data at all was available for only 303 patients (60.2%) (5).

In a 2018 survey of clinicians, 0% of survey respondents did reflex genetic testing of any kind for small bowel cancer (3)

1. Shaikh, T., Handorf, E. A., Meyer, J. E., Hall, M. J., & Esnaola, N. F. (2018). Mismatch Repair Deficiency Testing in Patients With Colorectal Cancer and Nonadherence to Testing Guidelines in Young Adults. *JAMA Oncol, 4*(2), e173580. doi:10.1001/jamaoncol.2017.3580
2. Jain, A., Shafer, L., Rothenmund, H., Kim, C. A., Samadder, J., Gupta, S., & Singh, H. (2019). Suboptimal Adherence in Clinical Practice to Guidelines Recommendation to Screen for Lynch Syndrome. *Dig Dis Sci, 64*(12), 3489-3501. doi:10.1007/s10620-019-05692-6
3. Pan, J. Y., Haile, R. W., Templeton, A., Macrae, F., Qin, F., Sundaram, V., & Ladabaum, U. (2018). Worldwide Practice Patterns in Lynch Syndrome Diagnosis and Management, Based on Data From the International Mismatch Repair Consortium. *Clin Gastroenterol Hepatol, 16*(12), 1901-1910 e1911.
4. Mathiak, M., Warneke, V. S., Behrens, H. M., Haag, J., Boger, C., Kruger, S., & Rocken, C. (2017). Clinicopathologic Characteristics of Microsatellite Instable Gastric Carcinomas Revisited: Urgent Need for Standardization. *Appl Immunohistochem Mol Morphol, 25*(1), 12-24. doi:10.1097/PAI.0000000000000264
5. Smyth, E. C., Wotherspoon, A., Peckitt, C., Gonzalez, D., Hulkki-Wilson, S., Eltahir, Z., . . . Cunningham, D. (2017). Mismatch Repair Deficiency, Microsatellite Instability, and Survival: An Exploratory Analysis of the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) Trial. *JAMA Oncol, 3*(9), 1197-1203. doi:10.1001/jamaoncol.2016.6762

[Response Ends]

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

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

[Response Begins]
N/A: race/ethnicity, gender, insurance and/or socioeconomic status, and disability data elements are not readily available in laboratory information systems and are therefore not captured in the Pathologists Quality Registry. It was determined by feasibility testing that these data elements would not be feasible to collect
[Response Ends]

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

[Response Begins]
Disparities exist in testing rates based on race, care setting and insurance status. A 2020 study found that testing rates are lower in community care hospitals than in academic hospitals, and White non-Hispanic patients are more likely to get testing than Black non-Hispanic patients. With respect to insurance: patients with no insurance were tested at 38%, private insurance, 47%. Medicare was only 36% and Medicaid 45% (1)

Similar studies from 2019 found lower rates of testing among racial and ethnic minorities as compared to non-Hispanic white patients, with only 6.0% of African American and 3.1% of Hispanic patients having genetic testing done, despite recommendations for universal screening for colorectal cancer (2-3)

- Froelich, W. (2020). Disparities in MSI/MMR Biomarker Testing for Colorectal Cancer. *Oncology Times, 42*(22). doi:10.1097/01.COT.0000723664.21999.cc
- Latham, A., Srinivasan, P., Kemel, Y., Shia, J., Bandlamudi, C., Mandelker, D., . . . Stadler, Z. K. (2019). Microsatellite Instability Is Associated With the Presence of Lynch Syndrome Pan-Cancer. *J Clin Oncol, 37*(4), 286-295. doi:10.1200/JCO.18.00283

- Llor, X. (2019). Lynch Syndrome: Widening the Net. *Gastroenterology, 157* (5), 1432-1434.

[Response Ends]

2. Scientific Acceptability of Measure Properties

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

---

sp.01. Provide the measure title.

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

[Response Begins]
Mismatch Repair (MMR) or Microsatellite Instability (MSI) Biomarker Testing Status in Colorectal Carcinoma, Endometrial, Gastroesophageal, or Small Bowel Carcinoma
[Response Ends]

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

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

[Response Begins]
Percentage of surgical pathology reports for primary colorectal, endometrial, gastroesophageal, or small bowel carcinoma, biopsy, or resection, that contain impression or conclusion of or recommendation for testing of mismatch repair (MMR) by immunohistochemistry (biomarkers MLH1, MSH2, MSH6, and PMS2), or microsatellite instability (MSI) by DNA-based testing status, or both
[Response Ends]

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

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

Please do not select:

- *Surgery: General*

[Response Begins]
Cancer: Colorectal
Cancer: Gynecologic
Cancer: Lung, Esophageal
[Response Ends]

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

[Response Begins]
Care Coordination
Other (specify)
CMS meaningful Measure Area: Transfer of Health Information and Interoperability
[Response Ends]

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

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

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

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]
Adults (Age >= 18)
[Response Ends]

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

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

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

Please do not select:

1. *Clinician: Clinician*
2. *Population: Population*

[Response Begins]
Clinician: Group/Practice
Clinician: Individual
[Response Ends]

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

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
[Response Begins]
Outpatient Services
[Response Ends]

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

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

[Response Begins]
https://documents.cap.org/documents/cap-33-MMR-MSI-for-checkpoint-inhibitor-therapy_v3.pdf
[Response Ends]

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

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.
[Response Begins]
Available in attached Excel or csv file
[Response Ends]

Attachment: MMR-MSI_2021DD.xlsx

sp.12. State the numerator.

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

DO NOT include the rationale for the measure.

[Response Begins]
Surgical pathology reports that contain impression or conclusion of or recommendation for testing of MMR by immunohistochemistry, MSI by DNA-based testing status, or both
[Response Ends]

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

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

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

[Response Begins]
This measure requires that immunohistochemistry (IHC) for the four MMR proteins (MLH1, MSH2, MSH6 and PMS2); or MSI by DNA-based testing; or both are addressed in the surgical pathology report for biopsy or resection specimens with primary or metastatic colorectal carcinoma and surgical pathology report for biopsy or resection specimens with primary or metastatic endometrial carcinoma are present. A short note can be made in the final report, such as or combination of:

- No loss of nuclear expression of MMR proteins
- Loss of nuclear expression of MMR proteins (intact expression)
- Microsatellite instability (MSI)
- Microsatellite instability high (MSI-H)
- Microsatellite instability low (MSI-L)
- Microsatellite stable (MSS)
- MMR, MSI, or both previously performed
- MMR, MSI, or both recommended
- MMR, MSI, or both cannot be determined

MMR/MSI status may be derived from either the primary or a reference laboratory, but the specific results (as noted above) need to be included within the final pathology report.
[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
All surgical pathology reports for primary colorectal, endometrial, gastroesophageal, or small bowel carcinoma, biopsy, or resection
[Response Ends]

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

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

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

[Response Begins]
CPT: 88305, 88307, 88309
AND
ICD-10:
•C18.0: Malignant neoplasm of cecum
•C18.2: Malignant neoplasm of ascending colon
•C18.3: Malignant neoplasm of hepatic flexure
•C18.4: Malignant neoplasm of transverse colon
•C18.5: Malignant neoplasm of splenic flexure
•C18.6: Malignant neoplasm of descending colon
•C18.7: Malignant neoplasm of sigmoid colon
•C18.8: Malignant neoplasm of overlapping sites of colon
•C18.9: Malignant neoplasm of colon, unspecified
•C19: Malignant neoplasm of rectosigmoid junction
•C20: Malignant neoplasm of rectum
•C54.1 Malignant neoplasm of endometrium
•C54.3 Malignant neoplasm of fundus uteri
•C54.8 Malignant neoplasm of overlapping sites of corpus uteri
•C54.9 Malignant neoplasm of corpus uteri, unspecified
•C55 Malignant neoplasm of uterus, unspecified

•C15.3: Malignant neoplasm of upper third of esophagus
•C15.4: Malignant neoplasm of middle third of esophagus
•C15.5: Malignant neoplasm of lower third of esophagus
•C15.8: Malignant neoplasm of overlapping sites of esophagus
•C15.9: Malignant neoplasm of esophagus, unspecified
•C16.0: Malignant neoplasm of cardia
•C16.1: Malignant neoplasm of fundus of stomach
•C16.2: Malignant neoplasm of body of stomach
•C16.3: Malignant neoplasm of pyloric antrum
•C16.4: Malignant neoplasm of pylorus
•C16.5: Malignant neoplasm of lesser curvature of stomach, unspecified
•C16.6: Malignant neoplasm of greater curvature of stomach, unspecified
•C16.8: Malignant neoplasm of overlapping sites of stomach
•C16.9: Malignant neoplasm of stomach, unspecified
•C17.0 Malignant neoplasm of duodenum
•C17.1 Malignant neoplasm of jejunum
•C17.2 Malignant neoplasm of ileum
•C17.3 Meckel's diverticulum, malignant
•C17.8 Malignant neoplasm of overlapping sites of small intestine

•C17.9 Malignant neoplasm of small intestine, unspecified
•C26.0 Malignant neoplasm of intestinal tract, part unspecified.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]
1) Patients with an existing diagnosis of Lynch Syndrome

2) Squamous cell carcinoma of the esophagus
[Response Ends]

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

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

[Response Begins]
Existing diagnosis of Lynch syndrome ICD-10 codes: ICD-10-CM Z15.0, Z15.04, Z15.09, Z80.0

[Response Ends]

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

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

[Response Begins]
N/A Measure is not risk stratified
[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.
[Response Begins]
No risk adjustment or risk stratification
[Response Ends]

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

Attachment: If available, please provide a sample report.
[Response Begins]
Rate/proportion
[Response Ends]

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

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

[Response Begins]

Better quality = Higher score

[Response Ends]

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

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

[Response Begins]

- Identify all surgical pathology reports for primary colorectal, endometrial, gastroesophageal, or small bowel carcinoma, including biopsy or resection specimens, and reports with other samples (i.e., resection specimen with lymph nodes)
- Remove cases positively identified as squamous cell carcinoma of the esophagus and cases where patient has an existing diagnosis of Lynch syndrome, regardless of how Lynch syndrome was diagnosed
- Identify cases from that population that have documentation of mismatch repair (MMR) testing by immunohistochemistry for all four of the MMR proteins (MLH1, MSH2, MSH6 and PMS2) OR microsatellite instability testing by DNA-based methods (PCR, next-generation or whole genome sequencing, multi-plex sequencing) OR documentation that such a test was recommended, ordered, or previously performed--the Met population
- From the remaining cases that do not have documentation of MMR/MSI testing status, remove any for which there is documentation of one of the following:
    - Insufficient tissue for testing or tissue is necrotic
    - Specimen contains metastatic carcinoma (not a primary neoplasm)
    - No residual tumor (post treatment)
    - Patient is not a candidate for checkpoint inhibitor therapy
- Remaining cases that do not have documentation of one of those medical reasons are considered Not Met for this measure

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

Electronic Health Data

Other (specify)

Laboratory Information Systems (LIS are not considered EHRs and are not eligible for CEHRT)

[Response Ends]

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

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

[Response Begins]
Currently in use in the Pathologists Quality Registry QCDR
[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]
No data collection instrument provided
[Response Ends]

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

• Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
• All required sections must be completed.
• For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
• If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
• An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
• Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
• For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.
2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.
2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.
2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
2b3. For outcome measures and other measures when indicated (e.g., resource use):
• an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
OR
• rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;
OR
there is evidence of overall less-than-optimal performance.
2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.
2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:
2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.
(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions
Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include but are not limited to inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
Examples of evidence that an exclusion distorts measure results include, but are not limited to frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
Risk factors that influence outcomes should not be specified as exclusions.
With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

2021 Submission:
Updated testing information here.

2018 Submission:
Testing from the previous submission here.

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

[Response Begins]

Electronic Health Data
Other (specify)
[Response Ends]

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

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

[Response Begins]
N/A, not an existing dataset
[Response Ends]

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

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

[Response Begins]
01-01-2019 - 12-31-2019
[Response Ends]

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

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

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

Please do not select:
- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Clinician: Group/Practice
Clinician: Individual
[Response Ends]

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

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

[Response Begins]
Fifty-one clinicians from academic and private practices submitted data for reliability testing. All submitted data was used in testing. Submitted data was deidentified information from manual chart abstraction
[Response Ends]

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

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

[Response Begins]
The population included 1282 cases of colorectal, endometrial, gastroesophageal, and small bowel cancer. Volunteer data abstractors were instructed to pull all cases for those diagnoses (per ICD-10 codes) for calendar 2019, no further sample was extracted. Per specifications, squamous cell carcinoma of the esophagus was removed from the denominator as an Exclusion. Only patients 18 and older were included. Laboratory information systems do not reliably capture race, ethnicity or most other demographic information including gender, so no separate analyses were performed based on those characteristics.
[Response Ends]

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

[Response Begins]
Different data was used for reliability and validity. Reliability testing was conducted using data from 51 pathologists. Validity testing was conducted as face validity with 40 subject matter experts including pathologists, gastroenterologists, and genetic counselors. There was not overlap between the two testing populations.
[Response Ends]

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

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

[Response Begins]
N/A; pathologists do not have access to this data
[Response Ends]

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

Reliability

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

Choose one or both levels.
[Response Begins]
Accountable Entity Level (e.g., signal-to-noise analysis)
[Response Ends]

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

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

[Response Begins]
Reliability of the measure score was calculated as the ratio of signal to noise using the methodological approaches from Rand[1]. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in provider performance and the noise is the total variability in measured performance. Testing was conducted using a beta binomial model with the beta distribution defined by alpha and beta to calculate within and between provider variance.

[1] Adams JL, The Reliability of Provider Profiling, Santa Monica, CA: RAND Corporation, 2009. https://www.rand.org/pubs/technical_reports/TR653.html
[Response Ends]

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

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

[Response Begins]
The mean reliability score was 0.96, with a standard deviation of 0.07, tenth percentile 0.91, ninetieth percentile 1.00. Assumed estimates based upon beta-binomial model distribution (using methods described by Adams in the 2009 Rand tutorial), alpha=0.2, beta=0.12. The raw performance scores had a mean of 60% with a standard deviation of 39 points, tenth percentile 0%, ninetieth percentile 100%.
[Response Ends]

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

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

[Response Begins]
A reliability equal to zero implies that all the variability in a measure is attributable to measurement error. A reliability equal to one implies that all the variability is attributable to real differences in physician performance. The reliability score for this measure is 0.96. Therefore, variability is almost entirely attributable to real differences in performance. Given that real variability was seen in the performance scores and confirmed by the volunteer clinicians, this is not unexpected and speaks to the potential for improvement in the measure.
[Response Ends]

Validity

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

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

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

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

[Response Begins]
Forty subject matter experts including pathologists, gastroenterologists and genetic counselors were provided the measure specifications in an online anonymous survey. They were then asked to rate their agreement with the following

statement: "The MMR/MSI Testing Status quality measure as described above will accurately distinguish between good and poor quality of care." on a scale of 1 to 4 where 1 is Strongly Disagree and 4 is Strongly Agree.
[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]
A score of 2.5 indicates face validity of the measure. For this measure, 22 experts answered 4 (Strongly Agree), 15 answered 3 (Agree), 1 answered 2 (Disagree) and 2 answered 1 (Strongly Disagree). The average score for the measure is 3.4
[Response Ends]

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

[Response Begins]
A score of 3.4 indicates that the measure has face validity in the population evaluated. Clinicians believe the measure as written will differentiate between good and poor care.

Participants were also given the opportunity to provide comments. In order to better understand the rationale of the 3 clinicians who disagreed, comments were evaluated. Interestingly, some clinicians supported a more stringent measure: rather than accept a statement recommending testing, some respondents suggested testing must be done in order to meet the measure. Other clinicians suggested that additional follow-up was needed for a positive test result, but this is not specified by the Guideline, so it is outside the scope of this measure. Of the clinicians who disagreed, one provided a comment indicating that he or she was unclear how the metric would be used and therefore lacked the context to evaluate it. This comment does not indicate a threat to face validity but instead a lack of clarity in the survey instructions; clinicians using this measure would understand that it is for MIPS.
[Response Ends]

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

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

[Response Begins]
Two methods were used for assessing meaningful differences in performance. First, volunteer data abstractors who pulled data for reliability testing were interviewed to determine whether performance gaps in reliability testing data represented real differences in testing rates, differences in documentation of testing, or issues in data abstraction such as the inability to access necessary data. For the fifty-one clinicians represented in the reliability data, abstractors confirmed that performance gaps represented real differences in testing rates for 33 clinicians and lack of documentation for 18 clinicians.

The second method for determining differences in performance is based on data in the Pathologists Quality Registry. In 2020, two of the diagnoses included in this measure were available in the Registry as standalone QCDR measures: MMR/MSI Testing Status in Colorectal Cancer and MMR/MSI Testing Status in Endometrial Cancer. Data submitted to CMS were evaluated to determine average performance score and thus meaningful differences in performance.
[Response Ends]

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

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

[Response Begins]
The average performance rate of clinicians who submitted data to CMS for the standalone colorectal cancer measure (see above) was 78.3%. Rates ranged from 40.35% to 100%. Twelve practices submitted data, representing 123 clinicians. The average performance rate of clinicians who submitted data to CMS for the standalone endometrial cancer measure was 77.4%. Rates ranged from 44.9% to 97.8%. Eight practices submitted data, representing 80 clinicians. Given that clinicians attested to CMS that these data are true, accurate, and complete, this represents real differences in performance.
[Response Ends]

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

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

[Response Begins]
Clinical guidelines recommend universal testing of colorectal and endometrial cancer samples for MMR/MSI. Data in the Registry indicates that universal testing is not occurring, a clinically significant finding. Given that the measures also allowed for pathologists to recommend testing, it is likely that rates of completed tests are actually lower than 78.3% and 77.4%, respectively. A benchmark has not been established for this measure, so performance cannot be graded in deciles. However, testing rates are meaningfully different than guideline recommendations.
[Response Ends]

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

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

[Response Begins]
Given that participants in the Pathologists Quality Registry can select any measures they want to report to CMS, it is possible that clinicians who chose not to enter data for the standalone colorectal cancer and endometrial cancer measures did so because they are not performing/recommending MMR/MSI tests. This cannot be determined from available data. To attempt to address this, the performance of the small number of clinicians who entered data into the Registry for the standalone colorectal cancer measure but did not submit it to CMS was assessed. Three practices entered data into the Registry but chose not to submit it to CMS. Of those, one practice was still refining natural language processing terms at the time of submission and therefore chose not to submit the data as they could not attest to its accuracy. The other two practices completed NLP. One had a performance score of 100%, the other had a performance score of 41.36%. It is possible that the latter practice chose not to submit data to CMS due to the low score, but this is unlikely for the former given that their score was perfect. Therefore, we cannot conclude that performance results are biased due to differences between responders and non-responders. As noted above, we cannot assess the performance of practices in the Registry who chose not to enter data on these measures.
[Response Ends]

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

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

[Response Begins]

Fifteen practices that submitted data to the Registry (including practices that did not report to CMS) on the standalone colorectal cancer. All practices submitted 100% of their data. Eight practices submitted data to the Registry for the standalone endometrial cancer measures (all reported data to CMS). One practice had a reporting rate of 92.8%. Therefore, of data entered into the Registry, frequency of missing data was very low and is unlikely to skew results.
[Response Ends]

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

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

[Response Begins]
As noted above, reporting of the measure is voluntary so we cannot fully assess the differences between responders and non-responders. However, within the population of clinicians who provided data, we do not have evidence of systematic missing data. No empirical analysis was conducted.
[Response Ends]

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

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

[Response Begins]
No, there is only one set of specifications for this measure
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
Yes, the measure uses exclusions.
[Response Ends]

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

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

[Response Begins]
Exclusions were tested as part of face validity testing. Alpha face validity testing was conducted during specification with a panel of clinical experts. At the time of assessment, only "Squamous cell carcinoma of the esophagus" was an Exclusion. This information is taken directly from the Guideline that is the basis of this measure. However, during alpha testing, it was determined that patients with an existing diagnosis of Lynch syndrome would not be a candidate for MMR/MSI testing. While MMR/MSI testing is one method to diagnose Lynch syndrome, other methods exist. Therefore the "patients with an existing diagnosis of Lynch syndrome" does not fully overlap with "testing previously performed", a Met condition for the measure. The relevant ICD-10 codes were included as an Exclusion.
[Response Ends]

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

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

[Response Begins]
Testing of Exclusions was included with face validity testing (see previous section). Forty subject matter experts reviewed the specifications including the Exclusions and asked to rate their agreement with the statement "The MMR/MSI Testing Status quality measure as described above will accurately distinguish between good and poor quality of care." on a scale of 1 (Strongly Disagree) to 4 (Strongly Agree). The average score of the measure was 3.4.

[Response Ends]

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

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

[Response Begins]
The score of 3.4 indicates the measure including the Exclusions has face validity. As noted above, one Exclusion was taken directly from the associated Guideline. The other was identified by clinical experts during measure development to prevent unnecessary testing of patients who have been diagnosed with Lynch syndrome by another method, as such testing would not add valuable information but would only increase expense and patient inconvenience.
[Response Ends]

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

[Response Begins]
No risk adjustment or stratification
[Response Ends]

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

[Response Begins]
N/A--not risk stratified. As this is a process measure, risk stratification was deemed inappropriate for this measure. Additionally, pathologists are non-patient-facing clinicians and as such do not have access to most patient data needed for any stratification or adjustment. This issue was identified while specifying the Exclusions and Exceptions and addressed via feasibility testing.
[Response Ends]

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

[Response Begins]
N/A--not an outcome measure
[Response Ends]

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

[Response Begins]
Other (specify)
N/A--not risk adjusted
[Response Ends]

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

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

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

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

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

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

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

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

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

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

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

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

[Response Begins]
N/A--not risk adjusted
[Response Ends]

3. Feasibility

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

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

[Response Begins]
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)
[Response Ends]

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

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.
[Response Begins]
ALL data elements are in defined fields in a combination of electronic sources
[Response Ends]

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

[Response Begins]
N/A, all data elements are available in a combination of electronic sources
[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]
N/A; data for this measure comes from Laboratory Information Systems, the pathology version of EHRs. Unlike EHRs, however, LISs are not eligible for CEHRT certification. Therefore, pathology clinical quality measures cannot currently be specified as eCQMs.
[Response Ends]

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

[Response Begins]
No feasibility issues identified so far. This measure is in use in 2021 in the Pathologists Quality Registry. As of October 2021, 201 clinicians from 16 practices were submitting data for this measure including 7 practices whose Laboratory Information Systems send data directly to the Registry without human intervention. In 2019 a version of this measure whose denominator was only colorectal carcinoma was in use in the Registry; 11 practices totaling 56 clinicians submitted data for the colorectal-cancer-only measure to CMS. In 2020, a version of the measure whose denominator was only endometrial carcinoma was also introduced. In 2020, 12 practices totaling 123 clinicians submitted the colorectal-carcinoma-only version and 8 practices totaling 80 clinicians submitted the endometrial-carcinoma-only version. Therefore, the measure has been operationalized as individual parts and as the current version in the Pathologists Quality Registry and can be directly electronically extracted from the medical record.

Feasibility testing using a score card was also conducted with pathologists, genetic counselors, and gastroenterologists. The only potential issue identified was that pathologists would not see specific documentation that a patient refused testing (identified in the specs as an example of a Denominator Exception, medical reason why testing status was not documented). Subsequent interviews confirmed that if a patient refused testing, no order would be sent to the pathologist, so the fact that the documentation would not appear in the LIS is moot.
[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]
N/A; no fees or licensing are required. No risk model, survey or instruments are needed.
[Response Ends]

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.

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

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

4a.01.

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

Name of program and sponsor

URL

Purpose

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

Level of measurement and setting

[Response Begins]
Payment Program
This measure is currently in use in the Merit-Based Incentive Payment System (MIPS; https://qpp.cms.gov/mips/reporting-options-overview).The MIPS program allows eligible clinicians to earn a payment adjustment for Medicare Part B covered professional services by demonstrating excellence in several performance categories including quality of care provided. This measure is one metric available to pathologists to demonstrate quality of care. It is currently in use as a QCDR measure, meaning it is only available to participants in the Pathologists Quality Registry. However, it has been submitted to the Measures Under Consideration list to be publicly available as a MIPS CQM (Clinical Quality Measure).

Since 2021 is the first year of use for this measure, data is still being collected. However, two previous versions of the measure were in use in 2019 and 2020: a standalone MMR/MSI testing in colorectal cancer measure, and a standalone MMR/MSI testing in endometrial cancer measure. These two were combined and additional diagnoses added to form the current version. In 2019, 11 practices totaling 56 clinicians submitted data for the colorectal-cancer-only measure to CMS. In 2020, the endometrial-carcinoma-only measure was introduced. In 2020, 12 practices totaling 123 clinicians submitted the colorectal-carcinoma-only version and 8 practices totaling 80 clinicians submitted the endometrial-carcinoma-only version.

As of October 2021, 16 practices are submitting data for the new combined measure. Although data are sent to the registry on a clinician-by-clinician basis (attributable to individual NPI-TIN combinations), in the past most practices have chosen to report at the practice level rather than the individual level. Each practice makes this decision at the time of submission to CMS (early 2022). Therefore, the level of measurement is currently unknown. These practices represent the following 11 states: California, Colorado, Florida, Idaho, Kentucky, Louisiana, Maine, Nebraska, New York, Pennsylvania, and Washington. Practices submitting data for this measure have as few as 3 cases or as many as 741 cases (to date). Data for a total of 2,680 patients have been submitted to the registry through October 2021.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Payment Program
[Response Ends]

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

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

[Response Begins]
N/A Measure is in use
[Response Ends]

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

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

[Response Begins]
N/A Measure is in use
[Response Ends]

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

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

[Response Begins]
Results in the form of performance rate (including initial population, denominator exclusions and exceptions, numerator/performance met, and performance not met) are available at any time to measured entities on the Pathologists Quality Registry dashboard, as are details about which cases fall into which category. The same information is available for the previous versions of the measure (standalone colorectal-cancer-only and endometrial-cancer-only) so practices can compare data to past performance.

During initial implementation, practices were provided a form to assist in determining which QCDR measures to use based on data they can access, procedures performed, and number of cases as well as interest in the topic area. Practices were given the chance to review the completed form with registry staff and/or to review the specifications. No results were available until practices submitted data to the registry; then results were available on the dashboard at any time.
[Response Ends]

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

[Response Begins]

Results in the form of performance rate (including initial population, denominator exclusions and exceptions, numerator/performance met, and performance not met) are available at any time to measured entities on the Pathologists Quality Registry dashboard. All participants receive emails at least once a quarter reminding them to check the dashboard. Practices who are submitting data via direction connection with the electronic medical record (the laboratory information system or LIS) have calls once a month with registry staff to review the dashboard and ensure that data is correct or determine why data was incorrectly categorized. Additionally, at any time, participants can "drill down" into their data and identify which cases fall into which category (Met, Not Met, Exclusion, etc.).

In addition, practices can schedule calls with registry staff at any time to ask questions or discuss performance on any measure. Registry staff also hold monthly "office hours" as an open forum for questions about the measure specifications.
[Response Ends]

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

[Response Begins]
Formal feedback about implementation has not been collected yet. Practices who are submitting data via direct connection with the laboratory information system report fewer than average issues in implementing this measure as compared with other QCDR measures in the registry. However, since data is still being collected, feedback on measure performance is not yet available.
[Response Ends]

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

[Response Begins]
N/A; 2021 is the first year of use of this measure, so feedback has not been collected. Feedback will be solicited from those who submit the measure to CMS in early 2022 as part of the regular registry assessment.
[Response Ends]

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

[Response Begins]
N/A; no other users available.
[Response Ends]

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

[Response Begins]
So far, no feedback has been received that would impact measure specifications. In response to implementation questions, registry staff hosted a virtual "office hours" to walk through measure specifications and data entry procedures. As the measure has been submitted to the MUC list in 2021, any feedback collected in early 2022 may or may not be implementable depending on MAP recommendations.
[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]
Because 2021 is the first year of use of this measure, no information about performance improvement or care improvement is available. However, a version of this measure which included only colorectal cancer was in use in 2019 and 2020. In 2019, the average performance rate was 64.8%, including data from 11 practices. In 2020, the average performance rate was 78.3%, including data from 12 practices. This increase in performance rate indicates that pathologists are more consistently documenting the status of MMR/MSI testing. This is not to say more MMR/MSI tests are being done; the measure is written to avoid incentivizing overuse. Instead, the measure incentivizes documentation of MMR/MSI status (including previously performed, recommended, ordered, etc.) in the pathology report. Based on these results, practices appear to have documented MMR/MSI testing status for colorectal cancer more frequently in 2020 vs 2019.

As of October 2021, 16 practices are using the combined MMR/MSI measure. This is more practices than either the standalone colorectal cancer or standalone endometrial cancer measure. This measure may therefore have greater utility to practices as a quality improvement tool.
[Response Ends]

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

[Response Begins]
N/A no unexpected findings identified so far
[Response Ends]

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

[Response Begins]
N/A no unexpected benefits realized so far.
[Response Ends]

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.

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

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

(Can search and select measures.)
[Response Begins]
[Response Ends]

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

(Can search and select measures.)
[Response Begins]
[Response Ends]

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

[Response Begins]
N/A. Searches in CMIT for "MMR", "MSI", "Mismatch repair" and/or "Microsatellite instability" yield no results. Similarly, searching for individual types of cancer (i.e., "endometrial carcinoma", "gastroesophageal carcinoma") yields no relevant results.
[Response Ends]

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

[Response Begins]
No
[Response Ends]

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

[Response Begins]
N/A, no related/competing measures
[Response Ends]

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

Provide analyses when possible.

[Response Begins]
N/A, no related/competing measures

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.: No appendix

Contact Information

Measure Steward (Intellectual Property Owner) : College of American Pathologists
Measure Steward Point of Contact: Skau, Colleen, cskau@cap.org
Measure Developer if different from Measure Steward: College of American Pathologists
Measure Developer Point(s) of Contact: Skau, Colleen, cskau@cap.org

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]
No appendix
[Response Ends]

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]
*Diana Marcella Cardona; Duke University--Chair of Quality and Clinical Data Registry Committee

Amarpreet Bhalla; Montefiore Medical Center

*Gregary Bocsi; University of Colorado School of Medicine--Vice-Chair for Quality Measures

Richard W Brown; Memorial Hermann

*Sarah M Eakin; University of Pittsburgh Medical Center

Yael K Heher; Harvard Medical School

*Richa Jain; University of Tennessee Health Science Center

Gordana Katava; Summit Medical Group

*Brent W Keenportz; CommonSpirit

Jennifer Laudadio; University of Arkansas for Medical Sciences

*Michelle L E Powers; Mercy Hospital Oklahoma City

C. Leilani Valdes; Northeast Pathology Group PA

Nicholas Bevins; University of California San Diego

All members reviewed measure specifications and provided input on data elements, rationale, and evidence. Members marked with asterisk (*) also participated in reliability testing.
[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]
2021
[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]
May 2021 for submission to MUC list
[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]
Yearly; reviewed each spring as part of annual measure updates
[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]
If the measure is recommended for use as a MIPS CQM as part of the 2021 MUC list, the measure will be reviewed in 2023 as part of the CMS annual measure maintenance cycle. If it is not recommended by the MAP in 2021, it will be reviewed/updated in spring of 2022 as part of the CAP annual measure updates.
[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]
N/A
[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]

[Response Ends]