NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 3059e

Corresponding Measures:

De.2. Measure Title: One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk

Co.1.1. Measure Steward: PCPI

De.3. Brief Description of Measure: Percentage of patients aged 18 years and older with one or more of the following: a history of injection drug use, receipt of a blood transfusion prior to 1992, receiving maintenance hemodialysis, OR birthdate in the years 1945–1965 who received one-time screening for hepatitis C virus (HCV) infection

1b.1. Developer Rationale: Of the estimated 3.5 million people living in the United States with the hepatitis C virus infection (HCV), only 50% have been tested for HCV and are aware of their status. Reported cases of HCV have increased (approximately 20% per year) between 2010-2016 which is partially due to improved case detection and more likely due to rising rates of injection drug use (1,2). Additionally, only one third have been referred for HCV care and only 5.6% receive recommended treatment (3). Studies indicate that even among high-risk patients for whom screening is recommended, only 49-75% are aware of their infection status (4,5,6). In a recent analysis of data from a national health survey, 67.9% of persons ever infected with HCV reported an exposure risk, (e.g., injection drug use, having sexual contact with suspected/confirmed hepatitis C patient), 2 weeks to 6 months prior to symptom onset, and the remaining 32.1% reported no known exposure risk (1). Current risk-based testing strategies have had limited success, as evidenced by the substantial number of HCV-infected persons who remain unaware of their infection. As a result, many do not receive needed care (e.g., education, counseling, and medical monitoring), and are not evaluated for treatment (5). HCV causes acute infection, which can be characterized by mild to severe illness but is usually asymptomatic. In approximately 75%-85% of persons, HCV persists as a chronic infection, placing infected persons at risk for liver cirrhosis, hepatocellular carcinoma (HCC), and extrahepatic complications that develop over the decades following onset of infection (6). HCV testing is the first step toward improving health outcomes for persons infected with HCV.

Citations

1. Centers for Disease Control and Prevention. Surveillance for viral hepatitis the United States, 2016. Available at: https://www.cdc.gov/hepatitis/statistics/2016surveillance/index.htm

2. Zibbell JE, Asher AK, Patel RC, Kupronis B, Iqbal K, Ward JW, Holtzman D. Increases in acute hepatitis C related infection related to a growing opioid epidemic and associated injection drug use, 2004-2014. Amer J Pub Health. 2018 Feb; 108(2):175-81.

3.. Holmberg SD, Spradling PR, Moorman AC, Denniston MM. Hepatitis C in the United States. N Engl J Med. 2013; 368:1859-1861. doi: 10.1056/NEJMp1302973

4.. Colvin HM, Mitchell AE. Hepatitis and Liver Cancer: A National Strategy for Prevention and Control of Hepatitis B and C. 2010. Available at: http://www.nap.edu/catalog/12793.html

5. Coffin PO, Reynolds A. Ending hepatitis C in the United States: the role of screening. Hepat Med. 2014:6 79–87.

6.. Hagan H, Campbell J, Thiede H, et al. Self-reported hepatitis C virus antibody status and risk behavior in young injectors. Public Health Rep 2006;121:710–9.

7. Alter MJ, Margolis HS, Krawczynski K, Judson, FN, Mares A, Alexander WJ, et al. The natural history of community-acquired hepatitis C in the United States. N Engl J Med 1992;327:1899–1905

S.4. Numerator Statement: Patients who received one-time screening for HCV infection

S.7. Denominator Statement: All patients aged 18 years and older who were seen twice for any visit or who had at least one preventive visit within the 12 month reporting period with one or more of the following: a history of injection drug use, receipt of a blood transfusion prior to 1992, receiving maintenance hemodialysis, OR birthdate in the years 1945–1965

S.10. Denominator Exclusions: Denominator Exclusions

Patients with a diagnosis of chronic hepatitis C

Denominator Exceptions

Documentation of medical reason(s) for not receiving one-time screening for HCV infection (eg, decompensated cirrhosis indicating advanced disease [ie, ascites, esophageal variceal bleeding, hepatic encephalopathy], hepatocellular carcinoma, waitlist for organ transplant, limited life expectancy, other medical reasons)

Documentation of patient reason(s) for not receiving one-time screening for HCV infection (eg, patient declined, other patient reasons)

De.1. Measure Type: Process

S.23. Data Source: Electronic Health Records

S.26. Level of Analysis: Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable.

Preliminary Analysis: New Measure (Previously Approved for Trial Use)

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 and Summary of prior review in 2015-2017

- Brief Summary: this is a measure of patients aged 18+ with a history of injection drug use, receipt of a blood transfusion prior to 1992, receiving maintenance hemodialysis, OR birthdate in the years 1945–1965 who received one-time screening for hepatitis C virus (HCV) infection.
- Developer provided a logic model depicting the relationship between one-time HCV screening, treating persons testing HCV-positive, and improved outcomes for persons infected with HCV.
- The developer provided a summary of the links between one-time hepatitis C screening and identifying hepatitis C positive which can lead to referrals/treatment and improved health outcomes.
- During its first review in 2015-2017 to the NQF Health and Well Being Committee, the developer previously presented guidelines from two societies (American Association for the Study of Liver Diseases [AASLD] and Infectious Disease Society of American [IDSA]) that recommend "persons should be screened for risk factors for HCV infection, and one-time testing should be performed for all persons with behaviors, exposures, and conditions associated with an increased risk of HCV infection" and high risk individuals and persons born between 1945 and 1965 without prior ascertainment of risk."
    - The developer provided an updated 2018 Guideline. AASLD-IDSA HCV Guidance Panel . Hepatitis C Guidance 2018 Update: AASLD-IDSA Recommendations for Testing, Managing, and Treating Hepatitis C Virus Infection. Clin Infect Dis. 2018 Oct 30;67(10):1477-1492. doi: 10.1093/cid/ciy585. URL: https://www.hcvguidelines.org/sites/default/files/full-guidance-pdf/HCVGuidance_May_24_2018b.pdf
    - Current 2018 Guideline cites the following recommendations:
        - "One-time HCV testing is recommended for persons born between 1945 and 1965* without prior ascertainment of risk." (Rating: Class I, Level B)
        - "Other persons should be screened for risk factors for HCV infection, and one-time testing should be performed for all persons with behaviors, exposures, and conditions associated with an increased risk of HCV infection." (Rating: Class I, Level B)
        - Class I recommendations refer to, "Conditions for which there is evidence and/or general agreement that a given diagnostic evaluation, procedure, or treatment is beneficial, useful, and effective."
        - Level B recommendation indicates that data are derived from a single randomized trial, nonrandomized studies, or equivalent
    - The developer summarized the Quality, Quantity, and Consistency of the body of evidence associated with the guideline

Changes to evidence from last review

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

☒ **The developer provided updated evidence for this measure (since its approval for trial use):**

Updates:

The developer provided the following additional clinical guidelines:

- Centers for Disease Control and Prevention (CDC). Recommendations for the Identification of Chronic Hepatitis C Virus Infection Among Persons Born During 1945–1965. MMWR 2012;61(No. RR-4): 1-36. URL: http://www.cdc.gov/hepatitis/hcv/guidelinesc.htm
    - "Adults born during 1945–1965 should receive one-time testing for HCV without prior ascertainment of HCV risk " (Strong Recommendation, Moderate Quality of Evidence)
    - Strength of recommendation and quality of evidence: Strong recommendation, moderate quality evidence. Definition bulleted below:
        - Clarity of balance between desirable and undesirable effects: Desirable effects clearly outweigh undesirable effects, or vice versa
        - Methodological Quality of Supporting Evidence: Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise or exceptionally strong evidence from unbiased observational studies
        - Implications: Recommendation can apply to most patients in most circumstances
    - The developer summarized the Quality, Quantity, and Consistency of the body of evidence associated with the guideline
- Moyer VA, on behalf of the U.S. Preventive Services Task Force. Screening for hepatitis C virus infection in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2013 Jun 25. URL: http://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/hepatitis-c-screening?ds=1&s=hepatitis%20c
    - "The USPSTF recommends screening for hepatitis C virus (HCV) infection in persons at high risk for infection. The USPSTF also recommends offering 1-time screening for HCV infection to adults born between 1945 and 1965." (Grade B recommendation) p. 350
    - The guideline also highlights a list of risk factors on p 351.
    - Grade B recommendations definition: "The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial."
    - The developer summarized the Quality, Quantity, and Consistency of the body of evidence associated with the guideline
- All 3 guidelines indicated a potential harm as anxiety on awaiting screening results and stigmatization for positive results.
- The developer cited a new study by Durham et al, 2016 on new hepatitis C treatment. No grading provided.

Questions for the Committee:

If the developer provided updated evidence for this measure since its approval for trial use:

- The evidence provided by the developer is updated, directionally the same compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?
- For structure, process, and intermediate outcome measures:
    - What is the relationship of this measure to patient outcomes?
    - How strong is the evidence for this relationship?
    - Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Process measure with systematic review (Box 3) ⍰⍰Summary of the QQC provided (Box 4) ⍰⍰Systematic review concludes moderate to high quality evidence.

The highest possible rating is "High" for Evidence

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.

- Note: developer did not perform separate analysis according to their specifications, which include multiple care settings.
- The developer provided current performance data on 3059e from Cerner EHR data from January 2017 through December 2017.
    - Based on the sample of 827 included providers, the mean performance rate is 0.20, the median performance rate is 0.17 and the mode is 0.14.
    - The standard deviation is 0.13.
    - The range of the performance rate is 0.98, with a minimum rate of 0.02 and a maximum rate of 1.0.
    - The interquartile range is 0.14 (0.25–0.11). Decile, Performance (1, 0.08; 2, 0.10; 3, 0.12; 4, 0.15; 5 ,0.17; 6 ,0.2; 7, 0.23; 8, 0.28; 9, 0.37; 10, 1.0).
- The developer provided current performance data on 3059e from a physician practice using the gGastro EHR for the time period January 2018 through December 2018.
    - Based on the sample of 180 included providers, the mean performance rate is 0.34, the median performance rate is 0.25 and the mode is 0.09.
    - The standard deviation is 0.26.
    - The range of the performance rate is 1.4, with a minimum rate of 0.01 and a maximum rate of 1.0.
    - The interquartile range is 0.31 (0.46–0.15). Decile, Performance (1, 0.08; 2, 0.13; 3, 0.18; 4, 0.23; 5, 0.25; 6,0.31; 7,0.4; 8,0.52; 9,0.78; 10, 1.0).
- The developer also cited literature on the performance gap/low rates of hepatitis c screening including:
    - National Health and Nutrition Examination Survey (NHANES) indicates that less than 50% of individuals living with HCV are currently aware of their status
    - Estimates from the NHANES indicate that approximately 45%-85% of individuals with chronic HCV infection are undiagnosed.

Disparities

- The developer did not provide disparities data on the measure. The developer predominantly provided literature on prevalence of hepatitis c in various population, but did cite one literature on disparities of screening of hepatitis C:
    - One study found that African Americans were more likely to be screened than Caucasians and men were more likely to be screened than women (Bourgi, 2016).

Questions for the Committee:

- Specific questions on information provided for gap in care.
- Is there a gap in care that warrants a national performance measure?
- If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

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?

- Evidence is strong and supports the use of the measure. The evidence directly applies to the process measure. Improving performance could result in improved pateint outcomes.
- The evidence is related to showing the link between one time screening and identifying hepatitis C positive which would lead to referrals. The outcome would mean that more people are identified and offered treatment for hepatitis C. This would then lead to improved health outcomes if the hepatitis C is treated instead of untreated with the potential to cause complications such as HCC, cirrhosis and morbidity.

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?

- strong evidence of a gap in care. Low screening rates. African Americans more likely to be screened than other races. Men more likely to be screened than women. Data demonstrating the dispartity was not provided by the measure developer. Cited studies such as CDC.
- There is a significant range in the performance rate, showing there is a definite gap in care. Overall screening rates are low. There is not evidence on disparities in all groups, although a single study was cited.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing

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

Reliability

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

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

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

eCQM Technical Advisor(s) review:

Submitted measure is an HQMF compliant eCQM	The submitted eCQM specifications follow the industry accepted format for eCQM (HL7 Health Quality Measures Format (HQMF)).
Documentation of HQMF,QDM, or CQL limitations	N/A – All components in the measure logic of the submitted eCQM are represented using the HQMF,QDM, or CQL standards;
Value Sets	The submitted eCQM specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results [from a simulated data set] demonstrating the measure logic can be interpreted precisely and unambiguously. – this includes 100% coverage of measured patient population testing with pass/fail test cases for each population;
Feasibility Testing	Number of data elements included in measure calculation: 23 Number of data elements scoring less than 3 on scorecard: 5 "Diagnosis: Injection Drug Use" • low scoring domains: standards ◦ comments on standards domain: standards currently available, but not widely adopted "Diagnosis: Limited Life Expectancy" • low scoring domains: availability, accuracy, standards, workflow ◦ comments on availability domain: Not defined in each system at this time ◦ comments on accuracy domain: The information may not be from the most authoritative source and/ or has a moderate likelihood of being correct. (e.g., self-report of a vaccination). ◦ comments on standards domain: Standards currently available, but not widely adopted ◦ comments on workflow domain: Additional time and effort over and above routine care is required but some perceived to be of benefit "Laboratory Test, Performed: HCV Antibody Test" • low scoring domains: standards ◦ comments on standards domain: Standards currently available, but not widely adopted "Laboratory Test, Performed not done: Medical Reason" • low scoring domains: standards, workflow ◦ comments on standards domain: Standards currently available, but not widely adopted ◦ comments on workflow domain: Additional time and effort over and above routine care is required but some perceived to be of benefit "Laboratory Test, Performed not done: Patient Reason" • low scoring domains: standards, workflow ◦ comments on standards domain: Standards currently available, but not widely adopted ◦ comments on workflow domain: Additional time and effort over and above routine care is required but some perceived to be of benefit

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

Evaluators: Primary Care and Chronic Illness project team staff

Link A (Project Team staff)

Evaluation of Reliability and Validity:

Reliability:

The developer conducted performance measure score reliability testing using two data samples listed below. Reliability testing was performed by using a beta-binomial model (i.e. signal to noise).

- Note: developer did not perform separate analysis according to their specifications, which include multiple care settings. Staff have rated this measure insufficient until developer can provide separate analyses of the reliability by care setting.
- Data 1: Cerner EHR data from a large non-profit health system with 23 hospitals, over 185 clinics, and over 2,000 providers with more than 130,000 admissions and 400,000 outpatient community care visits annually. From January 2017 through December 2017. This dataset reflects data from individual providers. 1,218 providers had all the required data elements and had at least 1 quality reporting event for a total of 115,053 quality events. The range of quality reporting events for 1,218 providers included is from 1 to 920. This sample appears to be potentially separatable into testing for care settings that match the specifications.
- Data 2: The data source is from a physician practice using the gGastro EHR. From January 2018 through December 2018. This dataset reflects data from individual providers. 189 providers had all the required data elements and had at least 1 quality reporting event for a total of 24,566 quality events. The range of quality reporting events for 189 providers included is from 1 to 730.

The developer had the following results for their two data samples.

- Developer has excluded providers with less than 10 quality reporting events. This is not included in the specifications. These must be included or stated as an exclusion criteria within the specifications.
- Data 1 (Cerner EHR): The reliability above 10 quality reporting events was 0.79. The reliability including providers with 1 or more quality reporting events is 0.68.
- Data 2 (gGastro EHR): The reliability above 10 quality reporting events was 0.96. The reliability including providers with 1 or more quality reporting events is 0.96.

For signal to noise, a reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high. Data 2 had very high reliability. Data 1 had acceptable reliability for providers with 10 or above quality reporting events, although below acceptable at 0.68 for providers with 1 or more quality reporting events.

Validity:

The developer submitted face validity testing for initial NQF submission. The developer noted they were unable to do any empirical validity since the measure is not in widespread use and noted this is initial endorsement so empirical validity not required for new NQF measures. The developer did face validity of the measure score with the expert panel for the measure. The expert panel included 17 members from the following specialty areas: community health, emergency medicine, gastroenterology, gerontology, hepatology, infectious disease, internal medicine, nursing, primary care, psychiatry, family medicine, medicine, and medical informatics.

They were asked to rate their agreement with the following statement:

- "The scores obtained from the measure as specified will accurately differentiate quality across providers."
- Scale 1-5, where 1=Strongly Disagree; 3=Neither Disagree nor Agree; 5=Strongly Agree

The results of the expert panel rating of the validity statement for Measure 3059e were as

follows:

- N = 17; Mean rating = 4.18 and 82.4% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.
- Several respondents felt that this measure would help to promote one-time screening and raise awareness of risk factor-based screening, but also highlighted that one-time screening is not enough for the high-risk populations.

The developer also conducted a denominator exception analysis:

The developer did a denominator exception analysis for frequency across providers for Data 1 (Cerner EHR) and Data 2 (gGastro EHR). See Denominator exceptions of this measure.

- Data 1 (Cerner EHR) : Amongst the 1,218 providers there were a total of 933 exceptions reported. The average number of exceptions per provider in this sample is 0.766. The proportion of exceptions to quality events is 0.008.
- Data 2 (gGastro EHR): Amongst the 189 providers there were a total of 1,280 exceptions reported. The average number of exceptions per provider in this sample is 6.80. The proportion of exceptions to quality events is 0.052.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, stratification approach, missing data, etc.)?
- The staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss validity?

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

- Measure developer has not separated analyses according to the measure's specifications; the measure must be tested by care setting.
- There two requirements for reliability must be met to be considered sufficient according to NQF critieria.

Evaluation A: Scientific Acceptability

Measure Number: 3059e

Measure Title: One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☒ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☒ **Clinician: Individual** ☐ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☒ **New** ☐ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.) **(previously approved for trial use measure)**

RELIABILITY: SPECIFICATIONS

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

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

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

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

   The Committee may want to discuss that the Denominator exceptions are appropriate, including patient reasons (i.e. patient declines).

   In addition, the denominator indicates it includes "patients who are seen twice for any visit or who had at least one preventive care." Committee may want to discuss why patient seen twice for any visit and not at one visit.

RELIABILITY: TESTING

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

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
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** N/A

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

   Submission document: Testing attachment, section 2a2.2

   - Note: developer did not perform separate analysis according to their specifications, which include multiple care settings. This has not been tested to specifications.
   - Developer excluded providers from the analysis with less than 10 events. This is not an exclusion criteria for the measure as part of the specifications. This measure has not been tested to specifications.
   - The developer conducted performance measure score reliability testing using two data samples listed below. Reliability testing was performed by using a beta-binomial model (i.e. signal to noise).
   - Data 1: Cerner EHR data from a large non-profit health system with 23 hospitals, over 185 clinics, and over 2,000 providers with more than 130,000 admissions and 400,000 outpatient community care

visits annually. From January 2017 through December 2017. This dataset reflects data from individual providers. 1,218 providers had all the required data elements and had at least 1 quality reporting event for a total of 115,053 quality events. The range of quality reporting events for 1,218 providers included is from 1 to 920.

- Data 2: The data source is from a physician practice using the gGastro EHR. From January 2018 through December 2018. This dataset reflects data from individual providers. 189 providers had all the required data elements and had at least 1 quality reporting event for a total of 24,566 quality events. The range of quality reporting events for 189 providers included is from 1 to 730.

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

   Submission document: Testing attachment, section 2a2.3

   The developer had the following results for their two data samples.

   - Data 1 (Cerner EHR): The reliability above 10 quality reporting events was 0.79. The reliability including providers with 1 or more quality reporting events is 0.68.
   - Data 2 (gGastro EHR): The reliability above 10 quality reporting events was 0.96. The reliability including providers with 1 or more quality reporting events is 0.96.

For signal to noise, a reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high. Data 2 had very high reliability. Data 1 had acceptable reliability for providers with 10 or above quality reporting events, although below acceptable at 0.68 for providers with 1 or more quality reporting events.

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 <u>all</u> testing results):

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

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

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

    ☒ **Insufficient** (NOTE: Should rate <u>INSUFFICIENT</u> 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.**

Data 2 had very high reliability; although Data 1 did have lower reliability with 0.68-0.79 range.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    The Committee may want to discuss that the Denominator exceptions are appropriate, including patient reasons (i.e. patient declines).

    The developer did a denominator exception analysis for frequency across providers for Data 1 (Cerner EHR) and Data 2 (gGastro EHR).

    - Data 1 (Cerner EHR) : Amongst the 1,218 providers there were a total of 933 exceptions reported. The average number of exceptions per provider in this sample is 0.766. The proportion of exceptions to quality events is 0.008.
    - Data 2 (gGastro EHR): Amongst the 189 providers there were a total of 1,280 exceptions reported. The average number of exceptions per provider in this sample is 6.80. The proportion of exceptions to quality events is 0.052.

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

    Submission document: Testing attachment, section 2b4.

The mean performance rates for this measure is extremely low at 0.20 for Data 1 (Cerner EHR) and 0.34 Data 2 (gGastro EHR).

- Data 1 (Cerner EHR)-The range of the performance rate is 1.0, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.20 (0.33–0.13).
- Data 2 (gGastro EHR)-The range of the performance rate is 0.99, with a minimum rate of 0.009 and a maximum rate of 1.00. The interquartile range is 0.34 (0.50–0.16).

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

    Submission document: Testing attachment, section 2b6.

    The developer notes there is no missing data and no missing data concerns.

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

    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

    16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

    16d.**Risk adjustment summary:** N/A

    16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

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

    16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)
☐ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

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

N/A

VALIDITY: TESTING

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

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

☒ Face validity

☐ Empirical validity testing of the measure score

☐ N/A (score-level testing not conducted)

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

Submission document: Testing attachment, section 2b2.2

The developer submitted face validity testing for initial NQF submission. The developer noted they were unable to do any empirical validity since the measure is not in widespread use and noted this is initial endorsement so empirical validity not required for new NQF measures. The developer did face validity of the measure score with the expert panel for the measure. The expert panel included 17 members from the following specialty areas: community health, emergency medicine, gastroenterology, gerontology, hepatology, infectious disease, internal medicine, nursing, primary care, psychiatry, family medicine, medicine, and medical informatics.

They were asked to rate their agreement with the following statement:

- "The scores obtained from the measure as specified will accurately differentiate quality across providers."
- Scale 1-5, where 1=Strongly Disagree; 3=Neither Disagree nor Agree; 5=Strongly Agree

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

Submission document: Testing attachment, section 2b2.3

The results of the expert panel rating of the validity statement for Measure 0359e were as follows:

- N = 17; Mean rating = 4.18 and 82.4% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.
- Several respondents felt that this measure would help to promote one-time screening and raise awareness of risk factor-based screening, but also highlighted that one-time screening is not enough for the high-risk populations.

21. **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)

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

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

☐ **Yes**

☐ **No**

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

23. **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.)

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

Level of analysis separate testing was not conducted according to specifications for individual clinicans and group/practice clinicians.

ADDITIONAL RECOMMENDATIONS

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

None

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?

- The developer did not submit data analysis on the selected settings. Would like to see testing in the various settings and with different populations. Does not state why the measure data was not anaysed for other settings that had been in the measure testing. Codes and descriptors are acceptable. Needs stronger testing.
- The reliability of Data2 was very high, but only acceptable for Data 1. The number of providers in the Data2 sample was fairly small, at 189 providers. Separate analysis was not performed. I suspect that it can be consistently implemented. I am not sure that patient declined needs to be a valid reason for exclusion.

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

- Developer did not follow their testing on the separate provider types as specified in their measure testing. More testing needed with different provider types.
- Data 1 has below acceptable reliability if all providers' data is included.

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

- Measure had face validity. More testing across various settings should be considered. However preliminary information indicates validity of the measure.
- Face validity seems appropriate given the newness of the measure. The face validity had a very high rate of agreement on the validity of the measure.

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?

- Should be tested by setting type, and with more providers/types. Current testing limits the validity.
- The range of performance supports that there is variation that warrants the measure.

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?

- Meassure developer excluded providers with less than 10. Not an exclusion based on measure specifications. Creates a bias of the results.
- The denominator exclusion for patient reasons (declines) appears to have led to large numbers of exclusions, especially in Data 2).

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 Specifications and Elements

- The measure is generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
- ALL data elements are in defined fields in a combination of electronic sources
- This measure is an eMeasure. 5 data elements scored low by NQF staff on the feasibility scorecard.

Data Collection Strategy

- The developer noted no data collection strategy issues.
- Commercial use of the measure would require a license agreement between user and PCPI Foundation.

The Health and Well Being Committee who previously reviewed this Approved for Trial use measure had discussed the one-time test and high risk behavior continuing and questioned the one-time only testing for hepatitis c. Also, that Committee noted that increase cost and lack of access to treatment (in particular to the Medicaid populations) remains a disincentive to test for the hepatitis C.

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?
- If an eCQM, does the eCQM Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?

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?

- Data elements exist to accurately report this measure.
- Age 1945-1965, and dialysis status are commonly used and easily capture on a problem list. The history of injection drug use has a clear ICD10, but is less commonly included on a problem list. It could be found in patient history, but will likely be recognized in different places. I think the blood transfusion prior to 1992 will be the most challenging. This is not routinely generated, and it would be difficult to capture the year of the transfusion in most data points.

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**

OR

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

Accountability program details

The registry version of the measure is in use in MIPS, but currently not this eMeasure version. The developer notes that they plan to submit this measure to the MIPS MUC List.

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 N/A -The measure is currently not used.

Additional Feedback: N/A -The measure is currently not used.

***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**

RATIONALE: The developer has plans to submit this measure to the MIPS MUC List.

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

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

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

Improvement results

No performance scores are available for the measure, at this time. The measure is currently not used.

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 measure is currently not in use in an accountability program.

Potential harms There are no harms identified by the developer. However, as the evidence indicated the potential harms as anxiety on awaiting screening results and stigmatization for positive results.

Additional Feedback: N/A

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?

- The measure is not currently being publicly reported.
- N/A

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.

- Performance results could increase testing and improve outcomes for patients through identification of patients with Hepatitis. Once testing has bene completed, and presuming it is done accurately, could drive improved patient outcomes.
- No concerns. This process measure would raise awareness of the screening measure and offer an opportunity to improve health outcomes should cases be treated.

Criterion 5: Related and Competing Measures

Related or competing measures
None
Harmonization
N/A

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 related or competing measures.
- N/A

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/12/2019

- **No NQF Members have submitted support/non-support choices as of this date.**

1. Evidence and Performance Gap – Importance to Measure and Report

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

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

NQF_evidence_attachment_3059e_04162019.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3059e

Measure Title: One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk

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

Date of Submission: 4/16/2019

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

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

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

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

☒ Process: Hepatitis C screening for patients at increased risk

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

At risk patients receive one-time HCV screening → Patients identified as HCV-positive are referred or receive treatment → Improved outcomes for persons infected with HCV

HCV is the most common chronic bloodborne infection in the United States (1). In the United States, an estimated 2.7 -3.6 million (1.0%-1.5%) persons in the general, non-institutionalized population are living with

hepatitis C virus (HCV) infection (2). Approximately 800,000 incarcerated, institutionalized, and homeless people are also infected with HCV (3). An estimated 41,200 persons were newly infected in 2016 (2). It is estimated that the average screening test for HCV costs between $45-$80 for uninsured individuals and less for those who are covered by insurance (4). Several analyses suggest that HCV testing is cost-effective and a reasonable strategy to identify asymptomatic cases, particularly in populations with a high prevalence of HCV. Rein and colleagues evaluated the cost-effectiveness of HCV screening in the primary care setting for the cohort born from 1945-1965. It was predicted that compared to the status quo, cohort screening would identify an additional 808,580 cases of HCV infection and prevent 82,000 HCV-related deaths. The cost per new case identified is $2,874 and $15,700 quality adjusted life years (QALY), assuming that the standard treatment of care is provided and $35,700 per QALY saved, assuming newer forms of treatment, including direct-acting antivirals, are used (5). Eckman and colleagues examined the cost-effectiveness of HCV screening and found that when used in populations with a prevalence of over .84%, it is cost-effective, particularly in populations with a higher prevalence of the infection (6). Additional studies have confirmed the cost-effectiveness of HCV testing linked to care and treatment (7). Given the current treatments available, it should be noted that 90% of cases can be cured, if identified (2).

1. Centers for Disease Control and Prevention. Viral Hepatitis - Hepatitis C Information. http://www.cdc.gov/hepatitis/HCV/index.htm Updated May 31, 2015. Accessed February 20, 2019.

2. Centers for Disease Control and Prevention. Surveillance for viral hepatitis the United States, 2016. Available at: https://www.cdc.gov/hepatitis/statistics/2016surveillance/index.htm

3. Edlin BR, Eckhardt BJ, Shu MA, et al. Toward a more accurate estimate of the prevalence of hepatitis C in the United States. Hepatology. 2015;62(5):1353-63.

4. Joshi SN. Hepatitis C screening. Ochsner J. 2014;14:664-668.

5. Rein DB, Smith BD, Wittenborn JS, et al. The cost-effectiveness of birth-cohort screening for hepatitis C antibody in U.S. primary care settings. Ann Intern Med. 2012;156(4):263-271.

6. Eckman MH, Talal AH, Gordon SC, et al. Cost-effectiveness of screening for chronic hepatitis C infection in the United States. Clin Infect Dis. 2013;56:1382-1393.

7. Rein DB, Wittenborn JS, Smith BD, Liffmann DK, Ward JW. The cost-effectiveness, health benefits, and financial costs of new antiviral treatments for hepatitis C virus. Clin Infect Dis. Clin Infect Dis. 201;61(2):157-68. doi: 10.1093/cid/civ220.

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

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☒ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	Title: HCV Guidance: Recommendations for Testing, Managing, and Treating Hepatitis C Author: American Association for the Study of Liver Diseases and Infectious Diseases Society of America Date: May 2018 Citation: AASLD-IDSA HCV Guidance Panel . Hepatitis C Guidance 2018 Update: AASLD-IDSA Recommendations for Testing, Managing, and Treating Hepatitis C Virus Infection. Clin Infect Dis. 2018 Oct 30;67(10):1477-1492. doi: 10.1093/cid/ciy585. URL: https://www.hcvguidelines.org/sites/default/files/full-guidance-pdf/HCVGuidance_May_24_2018b.pdf Title: Recommendations for the Identification of Chronic Hepatitis C Virus Infection Among Persons Born During 1945–1965 Author: Centers for Disease Control and Prevention (CDC) Date: 2012 Citation: Centers for Disease Control and Prevention (CDC). Recommendations for the Identification of Chronic Hepatitis C Virus Infection Among Persons Born During 1945–1965. MMWR 2012;61(No. RR-4): 1-36. URL: http://www.cdc.gov/hepatitis/hcv/guidelinesc.htm Title: Screening for hepatitis C virus infection in adults: U.S. Preventive Services Task Force recommendation statement Author: Moyer VA, on behalf of the U.S. Preventive Services Task Force Date: 2013 Citation: Moyer VA, on behalf of the U.S. Preventive Services Task Force. Screening for hepatitis C virus infection in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2013 Jun 25. URL: http://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/hepatitis-c-screening?ds=1&s=hepatitis%20c

Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** "One-time HCV testing is recommended for persons born between 1945 and 1965* without prior ascertainment of risk." (Rating: Class I, Level B) "Other persons should be screened for risk factors for HCV infection, and one-time testing should be performed for all persons with behaviors, exposures, and conditions associated with an increased risk of HCV infection. 1. Risk behaviors a. Injection drug use (current or ever, including those who injected once) b. Intranasal illicit drug use 2. Risk exposures a. Long-term hemodialysis (ever) b. Getting a tattoo in an unregulated setting c. Healthcare, emergency medical, and public safety workers after needle sticks, sharps, or mucosal exposures to HCV-infected blood d. Children born to HCV-infected women e. Prior recipients of transfusions or organ transplants, including persons who: i. Were notified that they received blood from a donor who later tested positive for HCV infection ii. Received a transfusion of blood or blood components, or underwent an organ transplant before July 1992 iii. Received clotting factor concentrates produced before 1987 f. Persons who were ever incarcerated 3. Other a. HIV infection b. Unexplained chronic liver disease and/or chronic hepatitis including elevated alanine aminotransferase levels c. Solid organ donors (deceased and living) *Regardless of country of birth" (Rating: Class I, Level B) **CDC Hepatitis C Screening Recommendations** • "Adults born during 1945–1965 should receive one-time testing for HCV without prior ascertainment of HCV risk (Strong Recommendation, Moderate Quality of Evidence), and • All persons identified with HCV infection should receive a brief alcohol screening and intervention as clinically indicated, followed by referral to appropriate care and treatment services for HCV infection and related conditions (Strong Recommendation, Moderate Quality of Evidence). Providers and patients can discuss HCV testing as part of an individual's preventive health care. For persons identified with HCV infection, CDC recommends that they receive appropriate care, including HCV-directed clinical preventive services (e.g., screening for alcohol use, hepatitis A and hepatitis B vaccination as appropriate, and medical monitoring of disease). Recommendations are available to guide treatment decisions. Treatment decisions should be made by the patient and provider after several factors are considered, including stage of disease, hepatitis C genotype, comorbidities, therapy-related adverse events, and benefits of treatment." p. 13 **USPSTF Recommendation** "The USPSTF recommends screening for hepatitis C virus (HCV) infection in persons at high risk for infection. The USPSTF also recommends offering 1-time screening for HCV infection to adults born between 1945 and 1965." (Grade B recommendation) p. 350 "Assessment of Risk

	The most important risk factor for HCV infection is past or current injection drug use. Another established risk factor for HCV infection is receipt of a blood transfusion before 1992. Because of the implementation of screening programs for donated blood, blood transfusions are no longer an important source of HCV infection. In contrast, 60% of new HCV infections occur in persons who report injection drug use within the past 6 months. Additional risk factors include long-term hemodialysis, being born to an HCV-infected mother, incarceration, intranasal drug use, getting an unregulated tattoo, and other percutaneous exposures (such as in health care workers or from having surgery before the implementation of universal precautions). Evidence on tattoos and other percutaneous exposures as risk factors for HCV infection is limited. The relative importance of these additional risk factors may differ on the basis of geographic location and other factors." p. 351

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** The AASLD and IDSA guideline recommendation has been assigned a Level B. Level B recommendation indicates that data are derived from a single randomized trial, nonrandomized studies, or equivalent **CDC Hepatitis C Screening Recommendations** The CDC guideline recommendation has been assigned a moderate evidence grade Strength of recommendation and quality of evidence: Strong recommendation, moderate quality evidence Clarity of balance between desirable and undesirable effects: Desirable effects clearly outweigh undesirable effects, or vice versa Methodological Quality of Supporting Evidence: Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise or exceptionally strong evidence from unbiased observational studies Implications: Recommendation can apply to most patients in most circumstances **USPSTF 2013 Recommendation** For each of the questions addressed in the review, the level of evidence was graded. According to the review, "We assessed the overall strength of each body of evidence as "high," "moderate," "low," or "insufficient" in accordance with the AHRQ 'Methods Guide for Effectiveness and Comparative Effectiveness Reviews' based on the quality of studies, consistency between studies, precision of estimates, and directness of evidence." Low quality studies are defined as "Low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate." Moderate quality studies are defined as, "Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate." **1. Does screening for HCV infection in nonpregnant adults without known abnormal liver enzymes reduce mortality and morbidity due to HCV infection, affect quality of life, or reduce incidence of HCV infection?** • No study compared clinical outcomes between individuals screened and not screened for HCV infection. No evidence. **2. What is the effectiveness of different risk- or prevalence-based methods for screening for HCV infection on clinical outcomes?** • No study compared clinical outcomes associated with different risk- or prevalence-based strategies for targeted HCV screening. No evidence. **3. What is the sensitivity and number needed to screen to identify 1 case of HCV infection of different risk- or prevalence-based methods for screening for HCV infection?** • Overall strength of evidence is low. 4**. What are the harms associated with screening for HCV infection, including diagnostic liver biopsies?** • Screening: overall strength of evidence: low. Liver biopsies: overall strength of evidence: moderate

Provide all other grades and definitions from the evidence grading system	**AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** Level A. Data derived from multiple randomized clinical trials, meta-analyses, or equivalent. B. Data derived from a single randomized trial, nonrandomized studies, or equivalent. C. Consensus opinion of experts, case studies, or standard of care. CDC Hepatitis C Screening Recommendations *See CDC grading framework table at the end of this document **USPSTF 2013 Recommendation** Attributes of high quality studies include, "High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect." Insufficient quality studies are defined as: "Evidence either is unavailable or does not permit a conclusion."
Grade assigned to the **recommendation** with definition of the grade	**AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** The AASLD and IDSA guideline recommendation has been assigned a Class I. Class I recommendations refer to, "Conditions for which there is evidence and/or general agreement that a given diagnostic evaluation, procedure, or treatment is beneficial, useful, and effective." **CDC Hepatitis C Screening Recommendations** The CDC guideline recommendation has been assigned a strong recommendation. The recommendations were developed using the GRADE methodology. The GRADE framework is an "approach that provides guidance and tools to define the research questions, conduct the systematic review, assess the overall quality of the evidence, and determine strength of the recommendation." A definition of the "strong recommendation" is provided below. Strength of recommendation and quality of evidence: Strong recommendation, moderate quality evidence Clarity of balance between desirable and undesirable effects: Desirable effects clearly outweigh undesirable effects, or vice versa Methodological Quality of Supporting Evidence: Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise or exceptionally strong evidence from unbiased observational studies Implications: Recommendation can apply to most patients in most circumstances **USPSTF Recommendation:** The USPSTF recommendation has been assigned a grade B. For grade B recommendations, "The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial."

Provide all other grades and definitions from the recommendation grading system	**AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** The AASLD and ISDA rating system used to rate the level of evidence and strength of the recommendation is based on scientific evidence and expert opinion and is an adaptation from the American College of Cardiology and American Heart Association Practice Guidelines. These classes and recommendations are summarized in the table below. Class I Evidence and/or general agreement that a given diagnostic evaluation, procedure, or treatment is beneficial, useful, and effective. II Conflicting evidence and/or a divergence of opinion about the usefulness and efficacy of a diagnostic evaluation, procedure, or treatment. IIa Weight of evidence and/or opinion is in favor of usefulness and efficacy. IIb Usefulness and efficacy are less well established by evidence and/or opinion. III Conditions for which there is evidence and/or general agreement that a diagnostic evaluation, procedure, or treatment is not useful and effective or if it in some cases may be harmful. **CDC Hepatitis C Screening Recommendations** *See CDC grading framework table at the end of this document **USPSTF Recommendation:** Grade A recommendations signify The USPSTF recommends the service. There is high certainty that the net benefit is substantial." Grade C recommendations signify "The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small." Grade D recommendation signify "The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits." I statements signify " The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined."

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	**Quantity:** **AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** 4 evidence-based guidelines, 1 systematic review, and 5 observational studies were cited in support of the recommendation statement to develop the recommendations and most relevant to the patient populations addressed in the measure. **CDC Hepatitis C Screening Recommendations** 10 randomized control trials, 35 observational studies, 4 systematic reviews, and 7 evidence based guidelines were cited in support of the recommendation statement to develop the recommendations and most relevant to the patient populations addressed in the measure. **USPSTF 2013 Recommendation** 16 observational studies were included to support the review and to develop the recommendations and most relevant to the patient populations addressed in the measure. **Quality:** **AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** Information regarding the overall quality of evidence across studies is not available. **CDC Hepatitis C Screening Recommendations** The quality of the evidence across studies is "moderate." **USPSTF 2013 Recommendation** 1. **Does screening for HCV infection in nonpregnant adults without known abnormal liver enzymes reduce mortality and morbidity due to HCV, affect quality of life, or reduce transmission of HCV?** No studies were available. 2. **What is the effectiveness of different risk- or prevalence-based methods for screening for HCV infection on clinical outcomes?** No studies were available. 3. **What is the sensitivity and number needed to screen to identify 1 case of HCV infection of different risk- or prevalence-based methods for screening for HCV infection?** Consistency was rated "high." 4. **What are the harms associated with screening for HCV infection, including diagnostic liver biopsies?** According to the review, consistency was rated as the following: "Screening: unable to assess (assessed different outcomes) Liver biopsies: moderate."

Estimates of benefit and consistency across studies	**AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** The guideline does not include an overall estimate of benefit from the body of evidence. However, the guideline does state, "HCV testing is recommended in select populations based on demography, prior exposures, high-risk behaviors, and medical conditions. Recommendations for testing are based on HCV prevalence in these populations, proven benefits of care and treatment in reducing the risk of hepatocellular carcinoma and all-cause mortality, and the potential public health benefit of reducing transmission through early treatment, viral clearance, and reduced risk behaviors." **CDC Hepatitis C Screening Recommendations** The guideline does not include an overall estimate of benefit from the body of evidence. However, the guideline does state, "Clinical preventive services, regular medical monitoring, and behavioral changes can improve health outcomes for persons with HCV infection." **USPSTF 2013 Recommendation** The review states, "As in the 2004 USPSTF review, we found no direct evidence on effects of HCV screening versus no screening on clinical outcomes, or on the comparison of clinical effects of alternative screening strategies. Retrospective studies found that screening strategies targeting multiple risk factors were associated with sensitivities exceeding 90% and numbers needed to screen to identify 1 case of HCV infection of less than 20. More narrowly targeted alternative screening strategies (such as screening only persons with a history of injection drug use) were associated with numbers needed to screen of less than 2, but they missed up to two thirds of infected patients." The review also states, "In the absence of direct evidence on clinical outcomes associated with screening, an indirect chain of evidence showing the availability of accurate diagnostic tests and effective treatments could link screening with improvements in clinical outcomes."
What harms were identified?	**CDC Hepatitis C Screening Recommendations** According to the guideline, although harms may entail patient worry such as anxieties and concern over waiting for testing results and concern of insurability, effective treatment is linked to reductions in greater liver-related morbidity and all-cause mortality. The guideline states: "A review of published and anecdotal evidence conducted in accordance with GRADE methodology indicated that the benefits of testing and treating persons with HCV infection were greater than the harms." **AASLD and IDSA Recommendations for Testing, Managing, and Treating Hepatitis C** The harms studied were not included in the guideline. However, it can be anticipated that harms of one-time HCV screening may include anxiety over waiting for test results and patient stigmatization. **USPSTF 2013 Recommendation** The systematic review states, "Although direct harms of screening appear minimal, harms such as labeling, anxiety, and stigmatization remain poorly studied and difficult to quantify."

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	1. **Citation:** Durham DP, Skirp LA, Bruce DB, et al. The impact of enhanced screening and treatment on hepatitis C in the United States. Clin Infect Dis. 2016;62(3):298-304. doi:10.1093/cid/civ894. 2. **Description:** Interferon-free direct-acting antivirals (DAAs), have been dramatically improving treatment options and outcomes for those undergoing HCV treatment. The authors of the article looked to quantify the benefit of using DAAs on HCV prevalence, morbidity and mortality through developing a transmission model that calibrated US epidemiological data stratified by age and injection drug use from 1992-2014. Durham and colleagues then projected the impact of enhanced HCV screening and use of DAAs on HCV infection and HCV-associated liver disease from 2015-2040. 3. **Results:** The model predicted that at current treatment rates, HCV prevalence in the US could fall by more than 80% in 2040 if DAAs were utilized. An additional 150,000 cases could be identified by increasing HCV screening among populations who do not inject drugs. If HCV screening is not expanded, at least 462,736 cases would remain untreated through 2040 among persons who inject drugs. 4. **Impact on conclusions:** Further opportunities in HCV screening exist, particularly for at-risk populations such as persons who inject drugs, and should be aggressively utilized in order to effectively lower HCV-related morbidity and mortality rates in the United States.

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

***Grades and definitions for CDC recommendations**

Strength of recommendation and quality of evidence	**Clarity of balance between desirable and undesirable effects**	**Methodological Quality of Supporting Evidence (examples)**	**Implications**
strong recommendation, high-quality evidence	Desirable effects clearly outweigh undesirable effects, or vice versa	Consistent evidence from well-performedRCTs or exceptionally strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research is unlikely tochange our confidence in the estimate of effect

Strength of recommendation and quality of evidence	Clarity of balance between desirable and undesirable effects	Methodological Quality of Supporting Evidence (examples)	Implications
strong recommendation, moderate quality evidence	Desirable effects clearly outweigh undesirable effects, or vice versa	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
strong recommendation, low-quality quality evidence	Desirable effects clearly outweigh undesirable effects, or vice versa	Evidence for at least one critical outcome from observational studies, RCTs with serious flaws or indirect evidence	Recommendation may change when higher quality evidence becomes available. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Strong recommendation very-low-quality evidence (very rarely applicable)	Desirable effects clearly outweigh undesirable effects, or vice versa	Evidence for at least one critical outcome from unsystematic clinical observations or very indirect evidence	Recommendation may change when higher quality evidence becomes available; any estimate of effect for at least one critical outcome is very uncertain
weak recommendation, high-quality evidence	Desirable effects closely balanced with undesirable effects	Consistent evidence from well-performedRCTs or exceptionally strong evidence from unbiased observational studies	The best action may differ depending on circumstances or patients or societal values. Further research is unlikely to change our confidence inthe estimate of effect.
weak recommendation, moderate-quality evidence	Desirable effects closely balanced with undesirable effects	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from unbiased observational studies	Alternative approaches likely to be better for some patients under some circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
weak recommendation, low-quality evidence	Uncertainty in the estimates of Desirable effects, harms, and burden; Desirable effects, harms, and burden may be closely balanced	Evidence for at least one critical outcome fromobservational studies, from RCTs with serious flaws or indirect evidence	Other alternatives may be equally reasonableFurther research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Strength of recommendation and quality of evidence	Clarity of balance between desirable and undesirable effects	Methodological Quality of Supporting Evidence (examples)	Implications
weak recommendation very low-quality evidence	Major uncertainty in the estimates of desirable effects, harms, and burden; Desirable effects may or may not be balanced with undesirable effects may be closely balanced	Evidence for at least one criticaloutcome from unsystematic clinical observations or very indirect evidence	Other alternatives may be equally reasonable. Any estimate of effect, for at least one critical outcome, is very uncertain

1b. Performance Gap

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

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

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

Of the estimated 3.5 million people living in the United States with the hepatitis C virus infection (HCV), only 50% have been tested for HCV and are aware of their status. Reported cases of HCV have increased (approximately 20% per year) between 2010-2016 which is partially due to improved case detection and more likely due to rising rates of injection drug use (1,2). Additionally, only one third have been referred for HCV care and only 5.6% receive recommended treatment (3). Studies indicate that even among high-risk patients for whom screening is recommended, only 49-75% are aware of their infection status (4,5,6). In a recent analysis of data from a national health survey, 67.9% of persons ever infected with HCV reported an exposure risk, (e.g., injection drug use, having sexual contact with suspected/confirmed hepatitis C patient), 2 weeks to 6 months prior to symptom onset, and the remaining 32.1% reported no known exposure risk (1). Current risk-based testing strategies have had limited success, as evidenced by the substantial number of HCV-infected persons who remain unaware of their infection. As a result, many do not receive needed care (e.g., education, counseling, and medical monitoring), and are not evaluated for treatment (5). HCV causes acute infection, which can be characterized by mild to severe illness but is usually asymptomatic. In approximately 75%-85% of persons, HCV persists as a chronic infection, placing infected persons at risk for liver cirrhosis, hepatocellular carcinoma (HCC), and extrahepatic complications that develop over the decades following onset of infection (6). HCV testing is the first step toward improving health outcomes for persons infected with HCV.

Citations

1. Centers for Disease Control and Prevention. Surveillance for viral hepatitis the United States, 2016. Available at:

https://www.cdc.gov/hepatitis/statistics/2016surveillance/index.htm

2. Zibbell JE, Asher AK, Patel RC, Kupronis B, Iqbal K, Ward JW, Holtzman D. Increases in acute hepatitis C related infection related to a growing opioid epidemic and associated injection drug use, 2004-2014. Amer J Pub Health. 2018 Feb; 108(2):175-81.

3.. Holmberg SD, Spradling PR, Moorman AC, Denniston MM. Hepatitis C in the United States. N Engl J Med. 2013; 368:1859-1861. doi: 10.1056/NEJMp1302973

4.. Colvin HM, Mitchell AE. Hepatitis and Liver Cancer: A National Strategy for Prevention and Control of Hepatitis B and C. 2010. Available at: http://www.nap.edu/catalog/12793.html

5. Coffin PO, Reynolds A. Ending hepatitis C in the United States: the role of screening. Hepat Med. 2014:6 79–87.

6.. Hagan H, Campbell J, Thiede H, et al. Self-reported hepatitis C virus antibody status and risk behavior in young injectors. Public Health Rep 2006;121:710–9.

7. Alter MJ, Margolis HS, Krawczynski K, Judson, FN, Mares A, Alexander WJ, et al. The natural history of community-acquired hepatitis C in the United States. N Engl J Med 1992;327:1899–1905

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Although not a requirement for initial endorsement, we have provided the following performance results from EHR measure 3059e to illustrate the current state of performance. There is no historical EHR data available at this time.

Data 1

The data source is Cerner EHR data from a large non-profit health system with 23 hospitals, over 185 clinics, and over 2,000 providers with more than 130,000 admissions and 400,000 outpatient community care visits annually. The data are for the time period January 2017 through December 2017. There were 113,303 patients included in this reliability testing and analysis. These were the patients that were associated with providers who had 10 or more patients eligible for this measure. Based on the sample of 827 included providers, the mean performance rate is 0.20, the median performance rate is 0.17 and the mode is 0.14. The standard deviation is 0.13. The range of the performance rate is 0.98, with a minimum rate of 0.02 and a maximum rate of 1.0. The interquartile range is 0.14 (0.25–0.11). Decile, Performance (1, 0.08; 2, 0.10; 3, 0.12; 4, 0.15; 5 ,0.17; 6 ,0.2; 7, 0.23; 8, 0.28; 9, 0.37; 10, 1.0).

Data 2

The data are for the time period January 2018 through December 2018. The data source is from a physician practice using the gGastro EHR. There were 24,520 patients included in this reliability testing and analysis. These were the patients that were associated with providers who had 10 or more patients eligible for this measure. Based on the sample of 180 included providers, the mean performance rate is 0.34, the median performance rate is 0.25 and the mode is 0.09. The standard deviation is 0.26. The range of the performance rate is 1.4, with a minimum rate of 0.01 and a maximum rate of 1.0. The interquartile range is 0.31 (0.46–0.15). Decile, Performance (1, 0.08; 2, 0.13; 3, 0.18; 4, 0.23; 5, 0.25; 6,0.31; 7,0.4; 8,0.52; 9,0.78; 10, 1.0).

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Data from the National Health and Nutrition Examination Survey (NHANES) indicates that less than 50% of individuals living with HCV are currently aware of their status (1). In addition, data from the CDC shows that a majority of patients (67.5%), reported an exposure risk at least 2 weeks to 6 months prior to symptom onset (2). Estimates from the NHANES indicate that approximately 45%-85% of individuals with chronic HCV infection are undiagnosed (3). A study at a large integrated health system found that among 444,594 patients that were seen between January 1, 2003 to August 1, 2012, 15.8% were screened for HCV (4). Of note in this study, higher risk groups were found to be screened more frequently, however rates are still suboptimal (4). A more recent study that focused on HCV screening at community health centers found that out of 60,772 eligible patients seen January 1, 2010, to December 31, 2013 8.3% had an HCV infection screen, as determined by presence of the HCV screen coded in the EHR (5). An evaluation of commercially insured patients between

2005-2014 showed that while rates of screening increased over time, rates remain overall low. Specifically, for those born between 1945-1965 screening rates increased from 1.71% in 2011 to 3.26% in 2014 (6).

Citations

1. Holmberg SD, Spradling PR, Moorman AC, Denniston MM. Hepatitis C in the United States. N Engl J Med. 2013; 368:1859-1861. doi: 10.1056/NEJMp1302973

2. Centers for Disease Control and Prevention. Surveillance for viral hepatitis the United States, 2016. Available at: https://www.cdc.gov/hepatitis/statistics/2016surveillance/index.htm

3. Ditah I, Dita F, Devaki P, et al. The changing epidemiology of hepatitis C virus infection in the United States: National health and nutrition examination survey 2001-2010. J Hepatol. 2014;60:691-698.

4. Linas BP, Hu H, Barter DM, Horberg M. Hepatitis C screening trends in a large integrated health system. The American Journal of Medicine. 2014. 127(5):398-405.

5. Cook N, Turse EP, Garcia AS, Hardigan P, Amofah SA. Hepatitis C virus infection screening within community health centers.

J Am Osteopath Assoc. 2016 Jan;116(1):6-11. doi: 10.7556/jaoa.2016.001.

6. Isenhour CJ, Hariri SH, Hales CM, Vellozzi CJ. Hepatitis C antibody testing in a commercially insured population, 2005-2014.

Am J Prev Med. 2017 May;52(5):625-631. doi: 10.1016/j.amepre.2016.12.016. Epub 2017 Feb 1.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*

No disparities data is available for the measure, at this time.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Overall, HCV risk-based screening strategies have at limited success as evidenced by the anywhere from 45-85% of people living with HCV being unaware of their infection. There are not only barriers to patients recalling risk behaviors, but also barriers with providers not being aware of the guidelines that support these screening strategies (7). Although the continued prevalence of HCV is problematic in communities across America, inequalities in disease prevalence, treatment, and outcomes make it a particularly important minority health issue. According to the CDC, American Indians and Alaska natives have the highest incidence of acute HCV cases and in 2016 had the highest rate of mortality from HCV when compared to other populations (1) While African Americans make up 12% of the U.S. population, they account for over 22% of chronic HCV cases (2). Additionally, African Americans diagnosed with HCV infection often have less desirable outcomes compared to white patients (3). In addition, chronic liver disease, often related to HCV infection, is a leading cause of death among African Americans aged 45-64 (2). A recent study in a large integrated healthcare system found that out of 40,561 patients eligible to be screened for HCV, 8657 patients (21.3%) were screened the using HCV antibody test and of these 109 (1.3%) tested positive (8). This study also found that African Americans were more likely to be screened than Caucasians and men were more likely to be screened than women (8). The “baby boomer” generation, or those born between 1945-1965, are six times more likely to have HCV infection than any other adult age group with an HCV prevalence of 3.25% (4). Other disparate groups include injection drug users. The most recent national surveys of injection drug users found that approximately one third of young drug users (aged 18-30 years) are HCV infected. Additionally, older and former injection drug users have a more common prevalence of HCV infection (70%-90%), primarily because of needle sharing during the 1970s and 1980s (5). Recent statewide surveillance data shows that HCV infections

have increased to an epidemic level in the central Appalachian region (Kentucky, Tennessee, Virginia and West Virginia), in individuals under the age of 30 primarily due to injection opioid use (6).

Citations

1. Centers for Disease Control and Prevention. Surveillance for viral hepatitis the United States, 2016. Available at:

https://www.cdc.gov/hepatitis/statistics/2016surveillance/index.htm

2. US Department of Health and Human Services. 2014-2016 Action Plan for the Prevention, Care, & Treatment of Viral Hepatitis. 2015. Available at: http://www.cdc.gov/hepatitis/HHS-ActionPlan.htm

3.. Webb BC. The "Secret" epidemic: Disparities in Hepatitis C Incidence, Treatment, and Outcomes. Prepared for the Joint Center for Political and Economic Studies. October 2010.

4. Denniston MM, Jiles RB, Drobeniuc J, et al. Chronic hepatitis C virus infection in the United States, National Health and Nutrition Examination Survey 2003 to 2010. Ann Intern Med. 2014 Mar 4;160(5):293-300. doi: 10.7326/M13-1133.

5.. Centers for Disease Control and Prevention. Hepatitis C FAQs for Health Professionals. http://www.cdc.gov/hepatitis/hcv/hcvfaq.htm#section1 Updated May 23, 2016. Accessed June 10, 2016.

6.. Centers for Disease Control and Prevention. Increases in hepatitis C virus infection related to injection drug use among persons aged =30 years - Kentucky, Tennessee, Virginia, and West Virginia, 2006–2012. MMWR. 2015;64(17):453-458.

7. Centers for Disease Control and Prevention. Recommendations for the Identification of Chronic Hepatitis C Virus Infection Among Persons Born During 1945-1965. MMWR 2012;61(No. RR-4):1-34.

8. Bourgi K, Brar I, Baker-Genaw K. Health disparities in hepatitis C screening and linkage to care in an integrated health system in southeast Michigan. PLOS One. Published: August 15, 2016 https://doi.org/10.1371/journal.pone.0161241.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;
  OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Patient/societal consequences of poor quality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

HCV is the most common chronic bloodborne infection in the United States (1). In the United States, an estimated 2.7 -3.6 million (1.0%-1.5%) persons in the general, non-institutionalized population are living with hepatitis C virus (HCV) infection (2). Approximately 800,000 incarcerated, institutionalized, and homeless people are also infected with HCV (3). An estimated 41,200 persons were newly infected in 2016 (2). It is estimated that the average screening test for HCV costs between $45-$80 for uninsured individuals and less for those who are covered by insurance (4). Several analyses suggest that HCV testing is cost-effective and a reasonable strategy to identify asymptomatic cases, particularly in populations with a high prevalence of HCV. Rein and colleagues evaluated the cost-effectiveness of HCV screening in the primary care setting for the

cohort born from 1945-1965. It was predicted that compared to the status quo, cohort screening would identify an additional 808,580 cases of HCV infection and prevent 82,000 HCV-related deaths. The cost per new case identified is $2,874 and $15,700 quality adjusted life years (QALY), assuming that the standard treatment of care is provided and $35,700 per QALY saved, assuming newer forms of treatment, including direct-acting antivirals, are used (5). Eckman and colleagues examined the cost-effectiveness of HCV screening and found that when used in populations with a prevalence of over .84%, it is cost-effective, particularly in populations with a higher prevalence of the infection (6). Additional studies have confirmed the cost-effectiveness of HCV testing linked to care and treatment (7). Given the current treatments available, it should be noted that 90% of cases can be cured, if identified (2).

1c.4. Citations for data demonstrating high priority provided in 1a.3

1. Centers for Disease Control and Prevention. Viral Hepatitis - Hepatitis C Information. http://www.cdc.gov/hepatitis/HCV/index.htm Updated May 31, 2015. Accessed February 20, 2019.

2 . Centers for Disease Control and Prevention. Surveillance for viral hepatitis the United States, 2016. Available at: https://www.cdc.gov/hepatitis/statistics/2016surveillance/index.htm

3. Edlin BR, Eckhardt BJ, Shu MA, et al. Toward a more accurate estimate of the prevalence of hepatitis C in the United States. Hepatology. 2015;62(5):1353-63.

4. Joshi SN. Hepatitis C screening. Ochsner J. 2014;14:664-668.

5. Rein DB, Smith BD, Wittenborn JS, et al. The cost-effectiveness of birth-cohort screening for hepatitis C antibody in U.S. primary care settings. Ann Intern Med. 2012;156(4):263-271.

6. Eckman MH, Talal AH, Gordon SC, et al. Cost-effectiveness of screening for chronic hepatitis C infection in the United States. Clin Infect Dis. 2013;56:1382-1393.

7. Rein DB, Wittenborn JS, Smith BD, Liffmann DK, Ward JW. The cost-effectiveness, health benefits, and financial costs of new antiviral treatments for hepatitis C virus. Clin Infect Dis. Clin Infect Dis. 201;61(2):157-68. doi: 10.1093/cid/civ220.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

Not applicable. Not a PRO-PM.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Infectious Diseases (ID), Liver : Viral Hepatitis

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

Disparities Sensitive, Screening

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

The measure specifications are included with this form. Additional measure details may be found at http://www.thepcpi.org/?page=PCPIMeasures

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

This is an eMeasure **Attachment:** OnetimeHCVscreenhighrisk_v5_6_Artifacts.zip

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

Attachment **Attachment:** HCVOnetimeScreenAtRisk_ValueSets_12182018-636876596793706335.xlsx

S.3. For endorsement maintenance, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.

Supporting guidelines and coding value sets included in the measure are reviewed on an annual basis. This annual review has resulted in minor changes to the value sets, to account for updates to the coding terminologies for existing data elements. Measure specifications are annually updated to align with any changes to the standards or tools used to support electronic measurement.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*

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

Patients who received one-time screening for HCV infection

S.5. Time Period for Data *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*

Time Period for Data Collection: the numerator quality action could happen any time before the end of the measurement period

S.6. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

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

NUMERATOR DEFINITION:

Screening for HCV Infection includes current or prior receipt of:

1) HCV antibody test
2) HCV RNA test
3) Recombinant immunoblot assay (RIBA) test (if performed at any time in the past)

NUMERATOR GUIDANCE:

This measure evaluates the proportion of at-risk patients who have received a one-time screening for Hepatitis C Virus (HCV). In order to meet the measure, the reporting provider must have the laboratory test result present in the patient´s medical record. On occasion, providers will view HCV screening results that were performed elsewhere and therefore the results are not present in the EHR in a structured format. To allow such tests to be applied to this measure, they should be entered into the EHR as a laboratory test in a manner

consistent with the EHR in use. If the specific LOINC code of the test is not known, the entry should use the more generic LOINC Panel code which is included in the HCV test value sets as outlined below:

If the provider does not know the exact HCV RNA test performed elsewhere, report the generic LOINC HCV RNA Panel code 75888-8, found in the value set titled, "HCV RNA Test".

If the provider does not know the exact HCV Antibody test performed elsewhere, report the generic LOINC HCV Ab Panel code, 75886-2, found in the value set titled, "HCV Antibody Test".

If the provider does not know the exact HCV RIBA test performed elsewhere, report the generic LOINC HCV RIBA Panel code, 75887-0, found in the value set, "HCV RIBA Test".

The following screening tests are included as allowable screening tests for HCV: HCV antibody test, HCV RNA test or RIBA test. The RIBA test qualifies as "one-time screening" if it was performed at some time in the past. Because RIBA is not a screening method currently used in clinical practice, it is not included as an option in the numerator logic for a screening that occurred during the measurement period.

HQMF eCQM developed and is attached to this submission in fields S.2a and S.2b.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*

All patients aged 18 years and older who were seen twice for any visit or who had at least one preventive visit within the 12 month reporting period with one or more of the following: a history of injection drug use, receipt of a blood transfusion prior to 1992, receiving maintenance hemodialysis, OR birthdate in the years 1945–1965

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

Time Period for Data Collection: 12 consecutive months

DENOMINATOR GUIDANCE

The start datetime stamp associated with the data element "Diagnosis: History of Blood Transfusion" should be the datetime of the transfusion event, and not a datetime stamp associated with the documentation action in order to satisfy the logic clause.

HQMF eCQM developed and is attached to this submission in fields S.2a and S.2b.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

Denominator Exclusions

Patients with a diagnosis of chronic hepatitis C

Denominator Exceptions

Documentation of medical reason(s) for not receiving one-time screening for HCV infection (eg, decompensated cirrhosis indicating advanced disease [ie, ascites, esophageal variceal bleeding, hepatic encephalopathy], hepatocellular carcinoma, waitlist for organ transplant, limited life expectancy, other medical reasons)

Documentation of patient reason(s) for not receiving one-time screening for HCV infection (eg, patient declined, other patient reasons)

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

Time Period for Data Collection: During the measurement period

The PCPI distinguishes between measure exceptions and measure exclusions. Exclusions arise when the intervention required by the numerator is not appropriate for a group of patients who are otherwise included in the initial patient or eligible population of a measure (ie, the denominator). Exclusions are absolute and are to be removed from the denominator of a measure and therefore clinical judgment does not enter the decision. For measure One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk, exclusions include Patients with a diagnosis of chronic hepatitis C. Exclusions, including applicable value sets, are included in the measure specifications.

Measure Exceptions

Exceptions are used to remove a patient from the denominator of a performance measure when the patient does not receive a therapy or service AND that therapy or service would not be appropriate due to patient-specific reasons. The patient would otherwise meet the denominator criteria. Exceptions are not absolute, and are based on clinical judgment, individual patient characteristics, or patient preferences. The PCPI exception methodology uses three categories of exception reasons for which a patient may be removed from the denominator of an individual measure. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For measure One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk, exceptions may include documentation of medical reason(s) for not receiving one-time screening for HCV infection, (eg, decompensated cirrhosis indicating advanced disease [ie, ascites, esophageal variceal bleeding, hepatic encephalopathy], hepatocellular carcinoma, waitlist for organ transplant, limited life expectancy, other medical reasons), or patient reason(s) (eg, patient declined, other patient reasons). Where examples of exceptions are included in the measure language, value sets for these examples are developed and are included in the eCQM. Although this methodology does not require the external reporting of more detailed exception data, the PCPI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The PCPI also advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

HQMF eCQM developed and is attached to this submission in fields S.2a and S.2b.

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)*

Consistent with CMS' Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF to standardize the collection of race and ethnicity data, we encourage the results of this measure to be stratified by race, ethnicity, administrative sex, and payer and have included these variables as recommended data elements to be collected.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)

No risk adjustment or risk stratification

If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*

No risk adjustment or risk stratification.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*

Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)

S.16. Type of score:

Rate/proportion

If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.18. Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*

To calculate performance rates:

1. Find the patients who meet the initial population (ie, the general group of patients that a set of performance measures is designed to address).
2. From the patients within the initial population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.
3. Find the patients who qualify for denominator exclusions and subtract from the denominator.
4. From the patients within the denominator (after denominator exclusions have been subtracted from the denominator), find the patients who meet the numerator criteria (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.
5. From the patients who did not meet the numerator criteria, determine if the provider has documented that the patient meets any criteria for exception when denominator exceptions have been specified [for this measure: medical reason(s) (eg, decompensated cirrhosis indicating advanced disease [ie, ascites, esophageal variceal bleeding, hepatic encephalopathy], hepatocellular carcinoma, waitlist for organ transplant, limited life expectancy, other medical reasons) or patient reason(s) (eg, patient declined, other patient reasons) for the patient not receiving one-time screening for HCV infection)]. If the patient meets any exception criteria, they should be removed from the denominator for performance calculation. -- Although the exception cases are removed from the denominator population for the performance calculation, the exception rate (ie, percentage of patients with valid exceptions) should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI.

If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No diagram provided

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF a PRO-PM, identify whether (and how) proxy responses are allowed.

Not applicable. The measure is not based on a sample.

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*

IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

Not applicable. The measure is not based on a survey.

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)

Required for Composites and PRO-PMs.

Patient eligibility is determined by a set of defined criteria relevant to a particular measure. If data required to determine patient eligibility are missing, those patients/cases would be ineligible for inclusion in the denominator and therefore the patient/case would be deleted.

If data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) are missing, this case would represent a quality failure.

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.24.

Electronic Health Records

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*

IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.

Not applicable.

S.25. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.26. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Clinician : Individual

S.27. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Home Care, Inpatient/Hospital, Other, Outpatient Services

If other: Domiciliary

S.28. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not applicable. The measure is not a composite.

2a. Reliability – See attached Measure Testing Submission Form

2b. Validity – See attached Measure Testing Submission Form

3059e_nqf_testing-attachment_7.1_-_Final_04162019-636918014001714521.docx

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

Measure Number (*if previously endorsed*)**:** 3059e
Measure Title: One-Time Screening for Hepatitis C Virus (HCV) for Patients at Risk
Date of Submission: 4/16/2019

Type of Measure:

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

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

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

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

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

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

Current testing data

Data 1

The data source is Cerner EHR data from a large non-profit health system with 23 hospitals, over 185 clinics, and over 2,000 providers with more than 130,000 admissions and 400,000 outpatient community care visits annually.

Data 2

The data source is from a physician practice using the gGastro EHR.

1.3. What are the dates of the data used in testing?

Current testing data

Data 1

The data are for the time period January 2017 through December 2017. Given the required conversion to ICD-10 in late 2015, the testing was completed on the ICD-10 specified measure.

Data 2

The data are for the time period January 2018 through December 2018. Given the required conversion to ICD-10 in late 2015, the testing was completed on the ICD-10 specified measure.

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Current testing data

Data 1

We received data from 1,218 providers reporting on this measure through the EHR in 2017. This dataset reflects data from individual providers and our analysis of the data is at the individual physician level. 1,218 providers had all the required data elements and had at least 1 quality reporting event for a total of 115,053 quality events. For this measure, 100 percent of providers are included in the analysis, and the average number of quality reporting events is 94. The range of quality reporting events for 1,218 providers included is from 1 to 920.

Data 2

We received data from 189 providers reporting on this measure through the EHR in 2018. This dataset reflects data from individual providers and our analysis of the data is at the individual physician level. 189 providers had all the required data elements and had at least 1 quality reporting event for a total of 24,566 quality events. For this measure, 100 percent of providers are included in the analysis, and the average number of quality reporting events is 129. The range of quality reporting events for 189 providers included is from 1 to 730.

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

Current testing data

Data 1

There were 115,053 quality reporting events included in this reliability testing and analysis. These were the events that were associated with providers who had 1 or more quality events eligible for this measure.

Data 2

There were 24,566 quality reporting events included in this reliability testing and analysis. These were the events that were associated with providers who had 1 or more quality events eligible for this measure.

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

Current testing Data 1 and Data 2

The same data samples were used for reliability testing and exceptions analysis.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime

rate) which do not have to be a proxy for patient-level data.

Current testing Data 1 and Data 2

Patient-level socio-demographic (SDS) variables were not captured as part of the testing.

2a2. RELIABILITY TESTING

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

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

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

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

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Data from both the 2017 and 2018 samples were tested using the same reliability testing method.

Reliability of the computed measure score was measured as the ratio of signal to noise. 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.

Reliability at the level of the specific provider is given by:

Reliability = Variance (provider-to-provider) / [Variance (provider-to-provider) + Variance (provider-specific-error]

Reliability is the ratio of the provider-to-provider variance divided by the sum of the provider-to-provider variance plus the error variance specific to a provider.

Reliability testing was performed by using a beta-binomial model. The beta-binomial model assumes the provider performance score is a binomial random variable conditional on the provider's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

Reliability is evaluated by averaging over provider specific reliabilities for all providers that submitted quality reporting events for the measure.

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 provider performance. A reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high. [1]

1. Adams JL, Mehrotra A, McGlynn EA, Estimating Reliability and Misclassification in Physician Profiling, Santa Monica, CA: RAND Corporation, 2010. www.rand.org/pubs/technical_reports/TR863. (Accessed on February 24, 2012.)

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

Current testing data

Data 1

The reliability above 10 quality reporting events was 0.79. The reliability including providers with 1 or more quality reporting events is 0.68.

Data 2

The reliability above 10 quality reporting events was 0.96. The reliability including providers with 1 or more quality reporting events is 0.96.

Table 1: EHR Reliability Results

	2017 Data 1	2018 Data 2
	Reliability	Reliability
1+ events	0.68	0.96
10+ events	0.79	0.96

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

Current testing data

Data 1

This measure has below acceptable reliability when evaluated at 1 or more quality reporting events and acceptable reliability when including providers with 10 or more quality reporting events.

Data 2

This measure has very high reliability when evaluated at 1 or more quality reporting events and very high reliability when including providers with 10 or more quality reporting events.

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

This measure is being submitted as a new measure for endorsement consideration. As such, the measure is not in widespread use and implementation has been limited, which make certain methods of validity testing (such as correlation analyses) challenging. As a result, the plans for validity testing this measure (which began in 2017) focused on face validity testing, based on previous communication with NQF staff that face validity

testing is sufficient for new measures. We understood this to be inclusive of submissions for new measures in the Spring 2019 cycle.

Current testing Data 1 and Data 2

Face Validity:

Input on the content validity of draft measures is obtained through soliciting comments from a panel of experts who were not involved in the development of the measure.

Face validity of the measure score as an indicator of quality was systematically assessed as follows:

After the measure was fully specified, the expert panel was asked to rate their agreement with the following statement: "The scores obtained from the measure as specified will accurately differentiate quality across providers".

Scale 1-5, where 1= Strongly Disagree; 3=Neither Agree nor Disagree; 5=Strongly Agree

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

The expert panel included 17 members from the following specialty areas: community health, emergency medicine, gastroenterology, gerontology, hepatology, infectious disease,

internal medicine, nursing, primary care, psychiatry, family medicine, medicine, and medical informatics.

1. Robert Wong, M.D. Alameda Health System Division of Gastroenterology and Hepatology, Alameda, CA (outpatient primary care) (Internal Medicine)
2. Marlene Moranino, RN, BSN, Community Health Center Association of Connecticut, Cheshire, CT (CHC) (Community Health)
3. Douglas White, M.D. Highland General Hospital Emergency Dept., Oakland, CA (Emergency Medicine)
4. Kristine Gonnella, MPH, National Nurse Led Care Consortium, Philadelphia, PA (Community Health)
5. Michelle Rose, MBA, Norton Healthcare, Louisville, KY (Outpatient, Inpatient, ED)
6. Lynn E. M. Taylor, M.D. Lifespan Health Care, Providence, RI (Primary Care, Infectious diseases)
7. Aaron Harris, M.D, CDC Atlanta, GA (Internal Medicine)
8. Andrew Hamilton RN, MS, Alliance Chicago, Chicago, IL (Community Health, Informatics, Nursing)
9. Barbara Resnick, PhD, RN, CRNP, FAAN, FAANP, University of Maryland, Baltimore, MD (Gerontology)
10. Andrew Saxon, M.D. VA Puget Sound Health Care System, Seattle, WA (Psychiatry)
11. Raymond Chung, M.D. Mass. General Hospital, Boston, MA (Gastroenterology)
12. Jason Wilson, M.D. USF, Tampa, FL (Internal Medicine)
13. Nancy Reau, M.D. Rush University Medical Center, Chicago, IL (Transplant Hepatology)
14. Grant Phillips, M.D. Maricopa County Correctional Health Services, Phoenix, AZ (Family Medicine)
15. Nancy Glick, M.D. ACCESS, Chicago, IL (Infectious Diseases)
16. Sarah Schillie, M.D. CDC, Atlanta, GA (Division of Viral Hepatitis)
17. Lance Stein, M.D. Piedmont, Atlanta, GA (Gastroenterology, Transplant Hepatology)

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

Current testing Data 1 and Data 2

The aforementioned expert panel was used to systematically assess face validity of the

measure. They were asked to rate their agreement with the following statement:

"The scores obtained from the measure as specified will accurately differentiate quality across

providers."

Scale 1-5, where 1=Strongly Disagree; 3=Neither Disagree nor Agree; 5=Strongly Agree

The results of the expert panel rating of the validity statement for Measure 0359e were as

follows:

N = 17; Mean rating = 4.18 and 82.4% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings

1 - 1 (Strongly Disagree)

2 - 1

3 - 1 (Neither Agree nor Disagree)

4 - 5

5 - 9 (Strongly Agree)

(a) Of 14 persons with either an 'agree' or 'strongly agree' response, 6 provided comments. Several respondents felt that this measure would help to promote one-time screening and raise awareness of risk factor-based screening, but also highlighted that one-time screening is not enough for the high-risk populations.

(b) Of the 2 persons with a 'disagree' or 'strongly disagree' response, 1 provided comments. That individual said that if patients are continuing to engage in high risk behavior that continues to put them at risk of acquiring HCV infection then one-time testing is not sufficient and annual testing may be more appropriate. The 1 person with a 'strongly disagree' response did not provide comments.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Current testing Data 1 and Data 2

The results of the expert panel rating of the validity statement for Measure 0359e were as

follows: N = 17; Mean rating = 4.18 and 82.4% of respondents either agree or strongly agree

that this measure can accurately distinguish good and poor quality. These results demonstrate that Measure 0359e is valid as specified.

2b2. EXCLUSIONS ANALYSIS

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

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Current testing Data 1 and Data 2

Exclusions include:

- Patients with a diagnosis of chronic hepatitis C

Exceptions include:

- Documentation of Medical reason for not receiving one-time screening for HCV infection (e.g., decompensated cirrhosis indicating advanced disease [i.e., ascites, esophageal variceal bleeding, hepatic encephalopathy], hepatocellular carcinoma, waitlist for organ transplant, limited life expectancy, other medical reasons)
- Documentation of patient reason(s) for not receiving one-time screening for HCV infection (e.g., patient declined, other patient reasons)

Exceptions were analyzed for frequency across providers.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Current testing data:

Data 1

Amongst the 1,218 providers there were a total of 933 exceptions reported. The average number of exceptions per provider in this sample is 0.766. The proportion of exceptions to quality events is 0.008.

Data 2

Amongst the 189 providers there were a total of 1,280 exceptions reported. The average number of exceptions per provider in this sample is 6.80. The proportion of exceptions to quality events is 0.052.

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exceptions are necessary to account for those situations when it is not medically appropriate to perform one-time screening for HCV infection. Exceptions are discretionary and the methodology used for measure exception categories are not uniformly relevant across all measures; for this measure, there is a clear rationale to permit an exception for a medical or patient reason. Rather than specifying an exhaustive list of explicit reasons for exception for each measure, the measure developer relies on clinicians to link the exception with a specific reason for the decision not to perform one-time screening for HCV infection.

Some have indicated concerns with exception reporting including the potential for providers to inappropriately exclude patients to enhance their performance statistics. Research has indicated that levels of exception reporting occur infrequently and are generally valid (Doran et al., 2008), (Kmetik et al., 2011). Furthermore, exception reporting has been found to have substantial benefits: "it is precise, it increases acceptance of [pay for performance] programs by providers, and it ameliorates perverse incentives to refuse care to "difficult" patients." (Doran et al., 2008).

Although this methodology does not require the external reporting of more detailed exception data, the measure developer recommends that providers document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. We also advocate for the systematic review and analysis of each provider's exceptions data to identify practice patterns and opportunities for quality improvement.

Without exceptions, the performance rate would not accurately reflect the true performance of that provider. This would result in an increase in performance failures and false negatives. The additional value of increased data collection of capturing an exception greatly outweighs the reporting burden.

References:

Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of pay for performance targets by English Physicians. New Engl J Med. 2008; 359: 274-84.

Kmetik KS, Otoole MF, Bossley H et al. Exceptions to Outpatient Quality Measures for Coronary Artery Disease in Electronic Health Records. Ann Intern Med. 2011;154:227-234.

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

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

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

Current testing Data 1 and Data 2

Not applicable

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

Current testing Data 1 and Data 2

Not applicable

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

Current testing Data 1 and Data 2

Not applicable

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

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

Current testing Data 1 and Data 2

Not applicable

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

Current testing Data 1 and Data 2

Not applicable

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Current testing Data 1 and Data 2

Not applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

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

Current testing Data 1 and Data 2

Not applicable

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

Current testing Data 1 and Data 2

Not applicable

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

Current testing Data 1 and Data 2

Not applicable

2b3.9. Results of Risk Stratification Analysis:

Current testing Data 1 and Data 2

Not applicable

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

Current testing Data 1 and Data 2

Not applicable

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

Current testing Data 1 and Data 2

Not applicable

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

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Current testing Data 1 and Data 2

Measures of central tendency, variability, and dispersion were calculated.

2b4.2. What were the statistical results from tesing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Current testing data

Data 1

Based on the sample of 1,218 included providers, the mean performance rate is 0.30, the median performance rate is 0.20 and the mode is 1.0. The standard deviation is 0.27. The range of the performance rate is 1.0, with a minimum rate of 0.00 and a maximum rate of 1.00. The interquartile range is 0.20 (0.33–0.13).

Data 2

Based on the sample of 189 included providers, the mean performance rate is 0.35, the median performance rate is 0.26 and the mode is 1.0. The standard deviation is 0.27. The range of the performance rate is 0.99, with a minimum rate of 0.009 and a maximum rate of 1.00. The interquartile range is 0.34 (0.50–0.16).

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

Current testing data

Data 1

The range of performance from 0.00 to 1.00 suggests there's clinically meaningful variation across providers' performance.

Data 2

The range of performance from 0.009 to 1.00 suggests there's clinically meaningful variation across providers' performance.

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

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

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

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Current testing Data 1 and Data 2

This test was not performed for this measure.

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

Current testing Data 1 and Data 2

This test was not performed for this measure.

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

Current testing Data 1 and Data 2

This test was not performed for this measure.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*

Current testing Data 1 and Data 2

The EHR datasets provided to us did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when creating the datasets in which case those values would not be counted towards measure performance. There is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various*

rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each)

Current testing Data 1 and Data 2

This test was not performed for this measure. There was no missing data.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Current testing Data 1 and Data 2

The EHR datasets provided to us did not contain missing data so this test was not performed. Nevertheless, missing data may have been rejected when creating the datasets in which case those values would not be counted towards measure performance. There is no indication that this missing data was systematic, thus their omission would lead to unbiased performance results.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in a combination of electronic sources

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.

Attachment **Attachment:** 3059e_Feasibility_Attachment_-_One_time_Screening.pdf

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For

eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

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

We have not identified any areas of concern or made any modifications as a result of testing of the measure in relation to data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, and other feasibility issues unless otherwise noted.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measures require a license agreement between the user and the PCPI® Foundation (PCPI®) or the American Medical Association (AMA). Neither the American Medical Association (AMA), nor the former AMA-convened Physician Consortium for Performance Improvement® (AMA-PCPI), nor PCPI, nor their members shall be responsible for any use of the Measures.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Planned	Current Use (for current use provide URL)
Public Reporting Payment Program Not in use	

4a.1. For each CURRENT use, checked above, provide:

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

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

The measure was included on the Measures Under Consideration (MUC) list as an eMeasure. While the measure was chosen by CMS for inclusion in PQRS 2016 and then rolled into the MIPS program, CMS chose to implement it as a registry measure. PCPI plans to resubmit this measure to the MUC list as an eMeasure with the hopes of it being implemented as such.

The PCPI strongly encourages the use of its measures in quality improvement and accountability initiatives and promotes their use in public reporting programs. Measures developed by the PCPI, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. As a measure developer, we work with measure implementers as opportunities arise to encourage and facilitate the integration of PCPI measures in their programs.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

The measure is a part of the Core Quality Measure Collaborative. The Collaborative intends to promote alignment and harmonization of measures across payers in the public and private sectors through core measure sets. CMS intends to include the core sets in proposed rules, where appropriate. Private payers will use a phased in approach to implementation of the core measure sets and may use them for negotiations between physicians and private payers. Other plans for implementation include regional and local quality improvement efforts using the core measure sets. The CQMC has reconvened in 2019 and we are continuing to participate in the process to become informed on the status of implementation of the core sets of measures. Please visit https://www.qualityforum.org/cqmc/ for additional information.

4b. Improvement

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

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

No performance scores are available for the measure, at this time

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Because a majority of individuals who are at-risk for HCV are not aware of their infection status, the implementation of this measure would lead to more individuals becoming aware of their infection status. Those who screen positive for HCV could seek access to follow-up care and treatment thereby improving their HCV-related outcomes.

4c. Unintended Consequences

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

4c.1. Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.

We are not aware of any unintended consequences at this time, but we take unintended consequences very seriously and therefore continuously monitor to identify actions that can be taken to mitigate them.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications completely harmonized?
5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

5b. Competing Measures

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

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): PCPI

Co.2 Point of Contact: PCPI, Measures, PCPImeasures@thepcpi.org, 312-224-6070-

Co.3 Measure Developer if different from Measure Steward: PCPI

Co.4 Point of Contact: Kerri, Fei, kerri.fei@thepcpi.org, 312-224-6070-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study are invited to participate as equal contributors to the measure development process. In addition, the PCPI strives to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

Work Group members:

John W. Ward, MD (Co-chair) (internal medicine)

John B. Wong, MD (Co-chair) (gastroenterology, hepatology, methodology)

Joel V. Brill, MD, AGAF, CHCQM (gastroenterology)

Roger Chou, MD (internal medicine, guideline experience)

Richard H. Davis, Jr., PA-C (physician assistant)

Yngve Falck-Ytter, MD, AGAF (gastroenterology/liver/hepatology)

Troy Fiesinger, MD, FAAFP (family medicine)

Marc G. Ghany, MD, MHSc (guideline experience/hepatology)

Marwan Haddad MD, MPH, AAHIVS (HIV/HCV, community health center)

Arthur Yu-shin Kim, MD (Infectious Diseases, HIV/HCV co-infection)

Pritha Kuchaculla (American Liver Foundation, patient advocacy)

Daniel B. Raymond (consumer/patient advocacy group)

Paola Ricci, MD (hepatology/gastroenterology)

Martha Shea, BSN, RN (American Liver Foundation, patient advocacy)

Jessica A. Shepherd, MD, MBA (OB/GYN)

Margaret C. Shuhart, MD, MS (hepatology/gastroenterology)

Amy Hirsch Shumaker, PharmD, BCPS (pharmacy, hepatology, infectious diseases)

Aynsley D. Smith, CNP, MPH (hepatology nurse practitioner)

Chris Taylor (patient advocacy/public health)

Glenn Treisman, MD, PhD (HIV and HCV psychiatrist)

Weifeng Weng, PhD (health services researcher/ABIM PIM development)

John Yao, MD, MPH, MBA, MPA, FACP (health plan representative)

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2013

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

Ad.4 What is your frequency for review/update of this measure? Supporting guidelines, specifications and coding for this measure are reviewed annually.

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

Ad.6 Copyright statement: Copyright 2019 PCPI(R) Foundation. All Rights Reserved.

The Measure is not a clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications.

The Measure, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measure require a license agreement between the user and the PCPI(R) Foundation (PCPI[R]) or the American Medical Association (AMA). Neither the AMA, nor the former AMA-convened Physician Consortium for Performance Improvement(R) (AMA-PCPI), nor PCPI, nor their members shall be responsible for any use of the Measure.

AMA and PCPI encourage use of the Measure by other health care professionals, where appropriate.

THE MEASURE AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. The AMA, the PCPI and its members and former members of the AMA-PCPI disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT[R]) or other coding contained in the specifications.

CPT(R) contained in the Measure specifications is copyright 2004-2018 American Medical Association. LOINC(R) is copyright 2004-2018 Regenstrief Institute, Inc. This material contains SNOMED Clinical Terms(R) (SNOMED CT[R]) copyright 2004-2018 International Health Terminology Standards Development Organisation. ICD-10 is copyright 2018 World Health Organization. All Rights Reserved.

Ad.7 Disclaimers: Please see the copyright statement above in AD.6 for disclaimer information.

Ad.8 Additional Information/Comments: