NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through National Quality Forum's (NQF) Consensus Development Process (CDP). The information submitted by the measure developers/stewards is included after the *Brief Measure Information* and *Preliminary Analysis* sections.

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

Brief Measure Information

NQF #: 0028

Corresponding Measures:

Measure Title: Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention

Measure Steward: National Committee for Quality Assurance

sp.02. Brief Description of Measure: Percentage of patients aged 18 years and older who were screened for tobacco use one or more times within the measurement period AND who received cessation counseling intervention on the date of the encounter or within the previous 12 months if identified as a tobacco user.

1b.01. Developer Rationale: This measure is intended to promote adult tobacco screening and tobacco cessation interventions for those who use tobacco products. There is good evidence that tobacco screening and brief cessation intervention (including counseling and/or pharmacotherapy) is successful in helping tobacco users quit. Tobacco users who are able to stop smoking lower their risk for heart disease, lung disease, and stroke.

sp.12. Numerator Statement:

Population 1: Patients who were screened for tobacco use at least once within the measurement period

Population 2: Patients who received tobacco cessation intervention

Population 3: Patients who were screened for tobacco use at least once within the measurement period AND who received tobacco cessation intervention if identified as a tobacco user

sp.14. Denominator Statement:

Population 1: All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period

Population 2: All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period who were screened for tobacco use and identified as a tobacco user

Population 3: All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period

sp.16. Denominator Exclusions:

Denominator Exclusions: not applicable. These are identified as denominator exceptions. We have defined the current methodology for distinguishing between denominator exclusions and denominator exceptions in question sp. 18. We did not receive testing information related to these exceptions.

Denominator Exceptions:

Population 1:

Documentation of medical reason(s) for not screening for tobacco use (eg, limited life expectancy, other medical reason)

Population 2:

Documentation of medical reason(s) for not providing tobacco cessation intervention (eg, limited life expectancy, other medical reason)

Population 3:

Documentation of medical reason(s) for not screening for tobacco use OR for not providing tobacco cessation intervention for patients identified as tobacco users (eg, limited life expectancy, other medical reason)

Measure Type: Process

sp.28. Data Source:

Claims

Registry Data

sp.07. Level of Analysis:

Clinician: Individual

IF Endorsement Maintenance – Original Endorsement Date: 08/10/2009

Most Recent Endorsement Date: 6/28/2017

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following description for this measure:

- This is a maintenance process measure at the clinician/individual level that determines the percentage of patients aged 18 years and older who were screened for tobacco use and who received cessation counseling intervention within 12 months if identified as a tobacco user.
- The developer provides a logic model that depicts cessation intervention for adults 18 and older who report tobacco use during screening is not only effective in helping users quit, but can also lower the risk for heart disease, stroke, and lung disease.

The developer provides the following evidence for this measure:

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

Summary of prior review in 2017

- The measure was previously based on 2015 U.S Preventive Services Task Force (USPSTF) recommendations for tobacco screening and cessation interventions.
- The developer summarized the Quality, Quantity, and Consistency of evidence to be high across measure components.
- The Standing Committee accepted the previous (2012) evaluation of evidence.

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:

- According to the developer, new evidence from systematic reviews completed between 2013 and 2022 as well as randomized and non-randomized trials show a link between receiving a cessation intervention and quitting among patients who report tobacco use.
- The developer used the 2021 Final Recommendation Statement from the USPSTF to gather the following evidence:
    - Clinicians should ask all adults, regardless of pregnancy status, about tobacco use.
    - The net benefit of behavioral and FDA-approved pharmacotherapy interventions, either alone or combined, for tobacco smoking cessation in nonpregnant adults is substantial.
    - The net benefit of behavioral interventions for tobacco smoking cessation in pregnant adults is substantial. Evidence regarding the benefits and harms of pharmacotherapy for tobacco cessation among pregnant adults is insufficient.
    - Evidence regarding the benefits and harms of using e-cigarettes for tobacco cessation among pregnant and nonpregnant adults is insufficient.

Exception to evidence

- N/A

Questions for the Standing Committee:

- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Standing Committee agree that there is no need for repeated discussion and a vote on evidence?*

- *What is the relationship between this measure and 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 is based on a systematic review (SR) of the evidence and the evidence is graded (Box 3) à Summary of QCC of the evidence provided (Box 4) à USPSTF Grade A with high Quantity, high Quality, high Consistency (Box 5a) à High

The highest possible rating is High.

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided 2020 (January 1-December 31, 2020) opt-in data from Merit-based Incentive Payment System which included 19,427 physicians and other clinicians (e.g. nurse practitioners, Physician Assistants).
    - The developer found a mean performance of 70.71 percent with a standand dievation of 29.93.
- The developer also looked at data from the 2015 National Health Interview Survey which analyzed the percentage of adults who attempted to quit and who received a cessation intervention. Data showed the following:
    - 68% of adults wanted to stop smoking
    - 55.4% made a quit attempt in the past year
    - 7.4% recently quit smoking
    - 57.2% were advised by a health professional to quit
    - 31.2% used medication and/or cessation counseling when trying to quit

Disparities

- The developer stated although this measure is included in federal reporting programs, disparities data has not yet been made available for analysis and reporting.
- The developer listed the following disparities data from the 2015 National Health Interview Survey:
    - Individuals 65 years and older (65.75 percent) and 45-64 years old (65.7 percent) were more likely to be advised to quit smoking compared to individuals 25-44 years old (49.8 percent) and 18-24 years old (44.4 percent).
    - White (60.2 percent) individuals were most likely to be advised to quit smoking compared to Hispanic (42.2 percent), Indian/Alaska Native (38.1 percent), and Asian (34.2 percent) individuals. Data on Black individuals were not included.
    - Smokers with any type of insurance (56.8-69.2 percent) were more likely to receive advice to quit than smokers who were uninsured (44.1 percent).

- Adults with a disability (71.8 percent) were more likely to be advised to quit smoking compared to adults without a disability (53.6 percent). Adults with serious psychological distress (70.2 percent) were more likely to be advised to quit smoking compared to adults without serious psychological distress (55.7 percent).

Questions for the Standing Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*

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

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

For maintenance measures—no change in emphasis—specifications should be evaluated the same as with new measures.

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

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

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

Specifications:

- Have the measure specifications changed since the last review? ☐ Yes ☒ No
- Measure specifications are clear and precise.

Reliability Testing:

- Did the developer conduct new reliability testing? ☒ Yes ☐ No
- Reliability testing conducted at the Accountable Entity Level:
  - The developer performed a signal-to-noise reliability test for each population (1: those screened for tobacco use, 2: those who received an intervention, and 3: those screened and who received an intervention). The data is from 19,427 individual physicians and other clinicians who opted-in to the Merit-based Incentive Payment System (MIPS).
  - For population 1, scores ranged from 0.888 to 1 with a mean reliability of 0.994. The 25th and 75th percentiles were 0.995 and 0.999 respectively, with a standard deviation of 0.012.
  - For population 2, scores ranged from 0.887 to 1 with a mean reliability of 0.992. The 25th and 75th percentiles were 0.993 and 0.998 respectively, with a standard deviation of 0.014.
  - For population 3, scores ranged from 0.896 to 1 with a mean reliability of 0.994. The 25th and 75th percentiles were 0.963 and 0.994 respectively, with a standard deviation of 0.023.
  - The developer attests the results indicate very good reliability, and that the variation is caused by real differences in performance across reporting entities.

Questions for the Standing Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*

Guidance From the Reliability Algorithm

Specifications are precise, unambiguous, and complete (Box 1) à Empirical reliability testing conducted with measure as specified (Box 2) à Testing conducted at the accountable entity level for each level of analysis (Box 4) à Reliability testing method was appropriate (Box 5) à Moderate level of certainty (Box 6b) à Moderate

The highest possible rating is MODERATE.

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

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

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

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

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

Validity Testing

- Did the developer conduct new validity testing? ☒ Yes ☐ No
- Validity testing conducted at the Accountable-Entity Level:
    - The developer performed a Pearson correlation test for construct validity to determine if the tobacco measure results correlate with another behavioral health screening measure, specifically the *Preventive Care and Screening: Unhealthy Alcohol Use: Screening and Brief Counseling*.
    - The developer notes a difference between the number of rates for the alcohol and tobacco measures. The alcohol measure has one rate assessing whether patients who were screened and identified as an unhealthy alcohol user received brief counseling, and the tobacco measure has three rates (1: those screened for tobacco use, 2: those who received an intervention, 3: those screened and who received an intervention). Each tobacco rate was assessed separately against the alcohol measure rate.
    - For population 1, the developer reports the rate is positively and moderately associated with the alcohol measure rate. The correlation coefficient was 0.461, and a p value <0.001.
    - For population 2, the developer reports the rate is positively and moderately associated with the alcohol measure rate. The correlation coefficient was 0.371, with a p value <0.001.
    - For population 3, the developer reports the rate is positively and moderately associated with the alcohol measure rate. The correlation coefficient was 0.434, with a p value <0.001.
    - The developer attests the tobacco measure performance is moderately associated with the alcohol screening measure. Therefore, the developer suggests that clinicians who perform well on one of these preventive behavioral health measures will likely perform well on the other.

Exclusions

- The measure does not use exclusions.

Risk Adjustment

- The measure is not risk-adjusted or stratified.

Meaningful Differences

- To demonstrate meaningful differences in performance, the developer calculated an IQR for each rate. An independent sample t-test was then calculated to determine if the differences were statistically significant.
    - There was an IQR of 49 for population 1, with a p value of <0.001.
    - There was an IQR of 47 for population 2, with a p value of <0.001.
    - There was an IQR of 61 for population 3, with a p value of <0.001.
- The developer attests the difference in performance is statistically significant for all three populations.

Missing Data

- The developer states they do not have information on the extent and distribution of missing data because this information is not collected by CMS.

Comparability

- The measure only uses one set of specifications for this measure.

Questions for the Standing Committee regarding validity:

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

Guidance From the Validity Algorithm

Potential threats to validity were empirically assessed (Box 1) → Empirical validity testing conducted using the measure as specified (Box 2) → Testing conducted at the accountable entity level for each level of analysis (Box 5) → Validity testing was appropriate (Box 6) → Moderate level of certainty (Box 7b) → Moderate

The highest possible rating is MODERATE.

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient5**

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- Data elements are collected and used by healthcare personnel.
- Claims data is caputured electronically with encounter codes for the denominator and CPT II codes for the numerator. However, the developer states registry implementation may vary.
- There is an eCQM version of this measure (CMS138), but it has been withdrawn from endorsement review. The developer had the following withdrawal request: "NCQA is not in a position to retest #0028e and therefore we think it best to remove the measure from the NQF portfolio."
- The developer has not identified feasibility issues related to data collection, availability of data, missing data, sampling, or patient confidentiality.

Questions for the Standing Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form (e.g., EHR or other electronic sources)?*
- *Is the data collection strategy ready to be put into operational use?*

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

Criterion 4: Use and Usability

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

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

4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and 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 they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.

Current uses of the measure

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

Accountability program details

- The measure is currently in the Quality Payment Program Merit-based Incentive Payment System, which is a payment program and also publicly shares submitted data and the Health Resources and Services Administration Uniform Data System, which provide a standardized reporting system for core set of information.
- The measure is also part of the Million Hearts Clinical Quality Measures, which are a focused set of high impact clinical quality measures for the ABCS (Aspirin when appropriate, Blood pressure control, Cholesterol management, and Smoking cessation) that are aligned across public and private national programs.

4a.2. Feedback on the measure provided 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; and (3) This feedback has been considered when changes are incorporated into the measure.

Feedback on the measure provided by those being measured or others

- CMS publishes measure performance results annually on its Physician Compare website. However, the developer notes that CMS does not provide information regarding feedback from measured entities on measure performance and implementation.

***Questions for the Standing Committee**:*

- *How can the performance results be used to further the goal of high quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

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

4b. Usability (4b1. Improvement; 4b2. Benefits of measure)

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

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

Improvement results

- The developer did not report any improvement results due to limited availability of QPP data.

4b2. Benefits versus harms. 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).

Unexpected findings (positive or negative) during implementation

- The developer did not report any unexpected findings.

Potential harms

- The developer did not report any potential harms.

***Questions for the Standing 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: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Criterion 5: Related and Competing Measures

Related Measures

- The developer states there are no endorsed related or competing measures but noted a related HEDIS measure titled Medical Assistance with Smoking and Tobacco Use Cessation [MSC]. The measure was endorsed by NQF under NQF #0027; however, endorsement was removed in 2020.

Harmonization

- This measure focuses on routine tobacco screening and tobacco cessation interventions with the intent of assessing clinician performance. NQF #0027 only assesses individuals who report smoking by capturing what percentage receive advice from a provider on quitting.

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

[Response Begins]

Yes

[Yes Please Explain]

New evidence continues to be published showing the link between receiving a cessation intervention and quitting among patients who report tobacco use. New research supports the ongoing practice of tobacco cessation interventions among all adults.

[Response Ends]

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

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]

Current Submission:

Adults age 18 and older >> screened for tobacco use >> Adults who report tobacco use receive a cessation intervention >> The intervention is effective in helping tobacco users quit >> Cessation of tobacco use lowers risk for heart and lung disease as well as stroke

Previous Submission:

This measure is intended to promote adult tobacco screening and tobacco cessation interventions for those who use tobacco products. There is good evidence that tobacco screening and brief cessation intervention (including counseling

and/or pharmacotherapy) is successful in helping tobacco users quit. Tobacco users who are able to stop smoking lower their risk for heart disease, lung disease, and stroke.

[Response Ends]

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

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

[Response Begins]

US Preventive Services Task Force Recommendation

[Response Ends]

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

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

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

[Response Begins]

Current Submission:

Title: Clinician Summary of USPSTF Recommendation "Interventions for Tobacco Smoking Cessation in Adults, including Pregnant Persons"

Author: UPSTF

Date: January 2021

Citation: U.S. Preventive Services Task Force. Tobacco Use in Adults and Pregnant Persons: Interventions. Jan 21. Retrieved from: https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/tobacco-use-in-adults-and-pregnant-women-counseling-and-interventions

Previous Submission:

1. Fiore MC, Jaen CR, Baker TB, et al. Treating tobacco use and dependence: 2008 update. Clinical practice guideline. Rockville, MD: U.S. Department of Health and Human Services. Public Health Service. May 2008.
2. U.S. Preventive Services Task Force. Counseling and interventions to prevent tobacco use and tobacco-caused disease in adults and pregnant women: U.S. Preventive Services Task Force reaffirmation recommendation statement. Ann Intern Med 2009 Apr 21;150(8):551-5.

[Response Ends]

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

[Response Begins]

Current Submission:

1. Nonpregnant adults: • Ask about tobacco use • Provide behavioral interventions and pharmacotherapy for cessation to those who use tobacco

2. Pregnant persons: • Ask about tobacco use • Provide behavioral interventions for cessation to those who use tobacco
3. Pregnant persons who use tobacco: • The evidence is insufficient to assess the balance of benefits and harms of pharmacotherapy for tobacco cessation
4. Adults and pregnant persons who use tobacco: • The evidence is insufficient to assess the balance of benefits and harms of using e-cigarettes for tobacco cessation. Clinicians should direct patients to other cessation interventions with proven effectiveness and established safety

Previous Submission:

PHS Guideline (1):

All patients should be asked if they use tobacco and should have their tobacco use status documented on a regular basis. Evidence has shown that clinic screening systems, such as expanding the vital signs to include tobacco use status or the use of other reminder systems such as chart stickers or computer prompts, significantly increase rates of clinician intervention. (Strength of Evidence = A)

All physicians should strongly advise every patient who smokes to quit because evidence shows that physician advice to quit smoking increases abstinence rates. (Strength of Evidence = A)

Minimal interventions lasting less than 3 minutes increase overall tobacco abstinence rates. Every tobacco user should be offered at least a minimal intervention, whether or not he or she is referred to an intensive intervention. (Strength of Evidence = A)

The combination of counseling and medication is more effective for smoking cessation than either medication or counseling alone. Therefore, whenever feasible and appropriate, both counseling and medication should be provided to patients trying to quit smoking. (Strength of Evidence = A)

Clinicians should encourage all patients attempting to quit to use effective medications for tobacco dependence treatment, except where contraindicated or for specific populations for which there is insufficient evidence of effectiveness (i.e., pregnant women, smokeless tobacco users, light smokers, and adolescents). (Strength of Evidence = A)

USPSTF Recommendation (2):

The USPSTF recommends that clinicians ask all adults about tobacco use and provide tobacco cessation interventions for those who use tobacco products. This is a grade A recommendation.

[Response Ends]

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

[Response Begins]

Current Submission:

1. Nonpregnant adults: • Ask about tobacco use • Provide behavioral interventions and pharmacotherapy for cessation to those who use tobacco
    a. Grade A
        i. The USPSTF recommends the service. There is high certainty that the net benefit is substantial.
2. Pregnant persons: • Ask about tobacco use • Provide behavioral interventions for cessation to those who use tobacco
    a. Grade A
        i. The USPSTF recommends the service. There is high certainty that the net benefit is substantial.
3. Pregnant persons who use tobacco: • The evidence is insufficient to assess the balance of benefits and harms of pharmacotherapy for tobacco cessation
    a. I Statement

        i. 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.

4. Adults and pregnant persons who use tobacco: • The evidence is insufficient to assess the balance of benefits and harms of using e-cigarettes for tobacco cessation. Clinicians should direct patients to other cessation interventions with proven effectiveness and established safety
    a. I Statement
        i. 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.

Previous Submission:

The quality of the body of evidence supporting each of the PHS guideline recommendations is summarized according to the strength of evidence ratings as "A." "A" evidence is described as "Multiple well-designed randomized clinical trials, directly relevant to the recommendation, yielded a consistent pattern of findings."

Additionally, the medication meta-analysis included predominantly studies with “self-selected” populations. In addition, in medication studies both experimental and control subjects in the studies typically received substantial counseling. Both of these factors tend to produce higher abstinence rates than typically are observed among self-quitters.

As a basis for their recommendations, the USPSTF reviewed new evidence in the PHS guideline.

[Response Ends]

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

[Response Begins]

Current Submission:

USPSTF does not grade each piece of evidence separately.

The USPSTF grades the quality of the overall evidence for a service on a 3-point scale (good, fair, poor):

- Good: Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes.
- Fair: Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies, generalizability to routine practice, or indirect nature of the evidence on health outcomes.
- Poor: Evidence is insufficient to assess the effects on health outcomes because of limited number or power of studies, important flaws in their design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes.

Previous Submission:

Every recommendation made by the PHS Panel bears a strength-of-evidence rating that indicates the quality and quantity of empirical support for the recommendation. Each recommendation and its strength of evidence reflects consensus of the Guideline Panel.

The three strength-of-evidence ratings are described as follows:

A. Multiple well-designed randomized clinical trials, directly relevant to the recommendation, yielded a consistent pattern of findings.
B. Some evidence from randomized clinical trials supported the recommendation, but the scientific support was not optimal. For instance, few randomized trials existed, the trials that did exist were somewhat inconsistent, or the trials were not directly relevant to the recommendation.
C. Reserved for important clinical situations in which the Panel achieved consensus on the recommendation in the absence of relevant randomized controlled trials.

[Response Ends]

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

[Response Begins]

Current Submission:

Grade A - The USPSTF recommends the service. There is high certainty that the net benefit is substantial.

I Statement - 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.

Previous Submission:

PHS: A; USPSTF does not separately grade the body of evidence

[Response Ends]

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

[Response Begins]

Current Submission:

- Grade A
    - **Definition**: The USPSTF recommends the service. There is high certainty that the net benefit is substantial.
    - **Suggestion for Practice**: Offer or provide this service.
- Grade B
    - **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.
    - **Suggestion for Practice**: Offer or provide this service.
- Grade C
    - **Definition**: 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.
    - **Suggestion for Practice**: Offer or provide this service for selected patients depending on individual circumstances.
- Grade D
    - **Definition**: 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.
    - **Suggestion for Practice**: Discourage the use of this service.
- I Statement
    - **Definition**: 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.
    - **Suggestion for Practice**: Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

Previous Submission:

Every recommendation made by the PHS Panel bears a strength-of-evidence rating that indicates the quality and quantity of empirical support for the recommendation. Each recommendation and its strength of evidence reflects consensus of the Guideline Panel.

The three strength-of-evidence ratings are described as follows:

A. Multiple well-designed randomized clinical trials, directly relevant to the recommendation, yielded a consistent pattern of findings.
B. Some evidence from randomized clinical trials supported the recommendation, but the scientific support was not optimal. For instance, few randomized trials existed, the trials that did exist were somewhat inconsistent, or the trials were not directly relevant to the recommendation.

C. Reserved for important clinical situations in which the Panel achieved consensus on the recommendation in the absence of relevant randomized controlled trials.

[Response Ends]

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

[Response Begins]

Current Submission:

The evidence published in the 2021 evidence review focused on the benefits and harms of behavioral and pharmacologic interventions for tobacco cessation in adults and pregnant adults. It relied on systematic reviews that occurred 2013 through 2022.

For behavioral interval counseling interventions among nonpregnant adults, 20 systematic reviews were included that covered approximately 830 randomized controlled trials (RCTs).

For pharmacotherapy for nonpregnant adults, 4 systematic reviews were included that covered 179 studies that were both non-randomized trials (NRT) and RCTs.

For combined pharmacotherapy and behavioral interventions among nonpregnant adults, 2 systematic reviews covered 117 studies. Most studies were NRT.

For behavioral counseling interventions among pregnant persons, 1 systematic review was conducted that included 97 studies. The studies were a combination of RCT and NRT.

For pharmacotherapy among pregnant persons, 1 review was conducted that included 5 NRT studies.

Previous Submission:

Since the measure essentially addresses three components (ie, (1) screening and cessation interventions comprising (2) brief counseling and/or (3) pharmacotherapy), the quantity of studies noted by the guideline are offered as they relate to each of the measure components.

For screening and assessment and its impact on clinical intervention, 9 studies met the selection criteria and were meta-analyzed.

For screening and assessment and its impact on tobacco cessation, 3 studies met the selection criteria and were meta analyzed.

For advice to quit smoking, 7 studies were included in the meta-analysis. For specific information about the intensity of the intervention, namely the efficacy of minimal counseling interventions lasting less than 3 minutes in comparison to low-intensity or high-intensity counseling interventions, 43 studies met the selection criteria for comparison across various lengths.

For combining counseling and medication, 18 studies met selection criteria.

For medication alone, a meta-analysis of 83 studies evaluated the effectiveness and abstinence rates for various medications and medication combinations compared to placebo at 6-months post-quit.

[Response Ends]

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

[Response Begins]

Current Submission:

The USPSTF found that the net benefit of providing behavioral interventions for adults and pregnant persons is substantial. Below is the language from the Task Force recommendation:

“The USPSTF concludes with high certainty that the net benefit of behavioral interventions and US Food and Drug Administration (FDA)–approved pharmacotherapy for tobacco smoking cessation, alone or combined, in nonpregnant adults who smoke is **substantial**.

The USPSTF concludes with high certainty that the net benefit of behavioral interventions for tobacco smoking cessation on perinatal outcomes and smoking cessation in pregnant persons is **substantial**.

The USPSTF concludes that the evidence on pharmacotherapy interventions for tobacco smoking cessation in pregnant persons is **insufficient** because few studies are available, and the balance of benefits and harms cannot be determined.

The USPSTF concludes that the evidence on the use of e-cigarettes for tobacco smoking cessation in adults, including pregnant persons, is **insufficient**, and the balance of benefits and harms cannot be determined. The USPSTF has identified the lack of well-designed, randomized clinical trials (RCTs) on e-cigarettes that report smoking abstinence or adverse events as a critical gap in the evidence.

See Table 1 for more information on the USPSTF recommendation rationale and assessment. For more details on the methods the USPSTF uses to determine net benefit, see the USPSTF Procedure Manual.”

Previous Submission:

For screening and assessment, the PHS panel looked at two different outcomes - the impact on clinical intervention and tobacco cessation. They concluded that "having a clinic system in place that identifies smokers increases rates of clinician intervention but does not, by itself, produce significantly higher rates of smoking cessation."

Results of the meta-analysis for advice to quit smoking show that brief physician advice significantly increases long-term smoking abstinence rates.

Results of the meta-analysis regarding the intensity of the counseling intervention revealed that all three session lengths (minimal counseling, low-intensity counseling, and higher intensity counseling) significantly increased abstinence rates over those produced by no-contact conditions.

However, there was a clear trend for abstinence rates to increase across these session lengths, with higher intensity counseling producing the highest rates.

For combining counseling and medication, the results of the meta-analysis indicate that providing counseling in addition to medication significantly enhances treatment outcomes.

For medication alone, the PHS Panel identified seven first-line (FDA-approved) medications (bupropion SR, nicotine gum, nicotine inhaler, nicotine lozenge, nicotine nasal spray, nicotine patch, and varenicline) and two second-line (non-FDA-approved for tobacco use treatment) medications (clonidine and nortriptyline) as being effective for treating smokers. Each has been documented to increase significantly rates of long-term smoking abstinence. These medications should be encouraged except where contraindicated or for specific populations for which there is insufficient evidence of effectiveness (i.e., pregnant women, smokeless tobacco users, light smokers, and adolescents).

As a basis for their recommendations, the USPSTF reviewed new evidence in the PHS guideline.

[Response Ends]

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

[Response Begins]

Current Submission:

“The USPSTF identified limited evidence on harms from behavioral counseling interventions for tobacco cessation. Three systematic reviews (1 on internet-based interventions, another on incentives, and 1 on hypnotherapy) did not find evidence of serious adverse events associated with interventions.”

[Response Ends]

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

[Response Begins]

There have been no studies published since the guideline that would significantly affect the findings.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

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

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

[Response Begins]

This measure is intended to promote adult tobacco screening and tobacco cessation interventions for those who use tobacco products. There is good evidence that tobacco screening and brief cessation intervention (including counseling and/or pharmacotherapy) is successful in helping tobacco users quit. Tobacco users who are able to stop smoking lower their risk for heart disease, lung disease, and stroke.

[Response Ends]

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

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

[Response Begins]

The data are from 19,427 physicians and other clinicians (e.g. nurse practitioners, Physician Assistants). The data were collected from individual providers who opt-in to MIPS. Due to the availability of QPP data, we are only able to provide performance scores for one year, 2020.

Measure Year: 2020 (January 1, 2020 through December 31, 2020)

Rate Distribution (%)

N	Mean	StdDev	Min	P10	P25	P50	P75	P90	Max
19,427	70.71%	29.93	0	23	49	81	98	100	100

[Response Ends]

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

[Response Begins]

A number of studies have shown steady rates of screening and cessation intervention during primary care and other office visits. The rate continues to show missed opportunities for cessation interventions.

Data collected via the National Health Interview Survey analyzed the percentage of adults who attempted to quit and who received a cessation intervention. Data from 2015 showed the following:

- "68% of adults wanted to stop smoking
- 55.4% made a past-year quit attempt
- 7.4% recently quit smoking
- 57.2% had been advised by a health professional to quit
- 31.2% used cessation counseling and/or medication when trying to quit"

1. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting Smoking Among Adults — United States, 2000–2015. MMWR Morb Mortal Wkly Rep 2017;65:1457–1464. DOI: http://dx.doi.org/10.15585/mmwr.mm6552a1.

[Response Ends]

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

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

[Response Begins]

While this measure is included in several federal reporting programs, those programs have not yet made disparities data available for us to analyze and report.

[Response Ends]

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

[Response Begins]

Data collected via the National Health Interview Survey showed that rates of screening and interventions varied by patients' race, age, insurance status, and disability status.

The stratification by age for smokers who received advice to quit are as follows:

- 18-24 years old – 44.4%
- 25-44 years old – 49.8%
- 45-64 years old – 65.7%
- 65+ years old – 65.75

The stratification by race for smokers who received advice to quit are as follows:

- Asian – 34.2%
- American Indian/Alaska Native – 38.1%
- Hispanic – 42.2%
- White – 60.2%

The stratification by insurance type for adults who received advice to quit are as follows:

- Uninsured– 44.1%
- Any type of insurance – 56.8%-69.2%

The stratification by disability status and psychological distress- for those who received advice to quit are as follows:

- Adults with a disability – 71.8%
- Adults with serious psychological distress – 70.2%
- Adults without disabilities – 53.6%
- Adults without serious psychological distress – 55.7%

1. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting Smoking Among Adults — United States, 2000–2015. MMWR Morb Mortal Wkly Rep 2017;65:1457–1464. DOI: http://dx.doi.org/10.15585/mmwr.mm6552a1.

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

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

spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

[Response Begins]

No

[Response Ends]

spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous NQF CDP review.

[Response Begins]

Not applicable. No material changes made.

[Response Ends]

sp.01. Provide the measure title.

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

[Response Begins]

Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention

[Response Ends]

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

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

[Response Begins]

Percentage of patients aged 18 years and older who were screened for tobacco use one or more times within the measurement period AND who received cessation counseling intervention on the date of the encounter or within the previous 12 months if identified as a tobacco user.

[Response Ends]

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

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

Please do not select:

- *Surgery: General*

[Response Begins]

Behavioral Health: Alcohol, Substance Use/Abuse

[Response Ends]

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

[Response Begins]

Primary Prevention: Tobacco Use

[Response Ends]

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

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

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

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Adults (Age >= 18)

[Response Ends]

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

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

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

Please do not select:

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

[Response Begins]

Clinician: Individual

[Response Ends]

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

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

[Response Begins]

Other

[Response Ends]

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

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

[Response Begins]

The measure specifications are included with this submission.

https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measures/2022_Measure_226_MIPSCQM.pdf

Claims: https://qpp.cms.gov/docs/QPP_quality_measure_specifications/Claims-Registry-Measures/2022_Measure_226_MedicarePartBClaims.pdf

[Response Ends]

sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

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

[Response Begins]

No data dictionary/code table – all information provided in the submission form

[Response Ends]

sp.13. State the numerator.

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

DO NOT include the rationale for the measure.

[Response Begins]

Population 1: Patients who were screened for tobacco use at least once within the measurement period

Population 2: Patients who received tobacco cessation intervention

Population 3: Patients who were screened for tobacco use at least once within the measurement period AND who received tobacco cessation intervention if identified as a tobacco user

[Response Ends]

sp.14. Provide details needed to calculate the numerator.

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

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

[Response Begins]

Time Period for Data Collection: At least once during the measurement period

Definitions:

Tobacco Use - Includes any type of tobacco

Tobacco Cessation Intervention - Includes brief counseling (3 minutes or less), and/or pharmacotherapy Note: For the purpose of this measure, brief counseling (e.g., minimal and intensive advice/counseling interventions conducted both in person and over the phone) qualifies for the numerator. Written self-help materials (e.g., brochures, pamphlets) and complementary/alternative therapies do not qualify for the numerator. Brief counseling also may be of longer duration or be performed more frequently, as evidence shows there is a dose-response relationship between the intensity of counseling provided (either length or frequency) and tobacco cessation rates (U.S. Preventive Services Task Force, 2015).

Numerator Note:

To satisfy the intent of this measure, a patient must have at least one tobacco use screening during the measurement period. If a patient has multiple tobacco use screenings during the measurement period, only the most recent screening, which has a documented status of tobacco user or tobacco non-user, will be used to satisfy the measure requirements.

If a patient uses any type of tobacco (i.e., smokes or uses smokeless tobacco), the expectation is that they should receive tobacco cessation intervention: either counseling and/or pharmacotherapy.

This measure defines tobacco cessation counseling as lasting 3 minutes or less. Services typically provided under CPT codes 99406 and 99407 satisfy the requirement of tobacco cessation intervention, as these services provide tobacco cessation counseling for 3- 10 minutes. If a patient received these types of services, submit G-code G9906 (for population criteria 1) and CPT Category II code 4004F (for population criteria 3).

Population 1:

Report quality data code:

G9902: Patient screened for tobacco use AND identified as a tobacco user

OR

G9903: Patient screened for tobacco use AND identified as a tobacco non-user

Population 2:

Report quality data code:

G9906: Patient identified as a tobacco user received tobacco cessation intervention (counseling and/or pharmacotherapy)

Population 3:

Report CPT Category II code:

4004F: Patient screened for tobacco use AND received tobacco cessation intervention (counseling, pharmacotherapy, or both), if identified as a tobacco user

OR

1036F: Current tobacco non-use

[Response Ends]

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

Population 1: All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period

Population 2: All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period who were screened for tobacco use and identified as a tobacco user

Population 3: All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period

[Response Ends]

sp.16. Provide details needed to calculate the denominator.

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

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

[Response Begins]

Time Period for Data Collection: At least once during the measurement period

Definitions:

Tobacco Use - Includes any type of tobacco

Denominator Note:

The denominator of submission criteria 2 is a subset of the resulting numerator for submission criteria 1, as submission criteria 2 is limited to assessing if patients identified as tobacco users received an appropriate tobacco cessation intervention. For all patients, submission criteria 1 and 3 are applicable, but submission criteria 2 will only be applicable for those patients who are identified as tobacco users. Therefore, data for every patient that meets the age and encounter requirements will only be submitted for submission criteria 1 and 3, whereas data submitted for submission criteria 2 will be for a subset of patients who meet the age and encounter requirements, as the denominator has been further limited to those who were identified as tobacco users.

Population 1:

Patients aged >= 18 years on date of encounter

AND

At least two patient encounters during the performance period (CPT): 90791, 90792, 90832, 90834, 90837, 90845, 92002, 92004, 92012, 92014, 92521, 92522, 92523, 92524, 92540, 92557, 92625, 96150, 96151, 96152, 97165, 97166, 97167, 97168, 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350

WITHOUT

Telehealth Modifier: GQ, GT, 95, POS 02

OR

At least one preventive encounter during the performance period (CPT or HCPCS): 99385, 99386, 99387, 99395, 99396, 99397, 99401, 99402, 99403, 99404, 99411, 99412, 99429, G0438, G0439

WITHOUT

Telehealth Modifier: GQ, GT, 95, POS 02

Population 2:

Patients aged >= 18 years on date of encounter

AND

All eligible instances when (G9902) Patient screened for tobacco use AND identified as a tobacco user that are utilized in submission of Performance Met Patient Screened for Tobacco Use, Identified as a Tobacco User in the numerator for population one

AND

At least two patient encounters during the performance period (CPT): 90791, 90792, 90832, 90834, 90837, 90845, 92002, 92004, 92012, 92014, 92521, 92522, 92523, 92524, 92540, 92557, 92625, 96150, 96151, 96152, 97165, 97166, 97167,

97168, 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350

WITHOUT

Telehealth Modifier: GQ, GT, 95, POS 02

OR

At least one preventive encounter during the performance period (CPT or HCPCS): 99385, 99386, 99387, 99395, 99396, 99397, 99401, 99402, 99403, 99404, 99411, 99412, 99429, G0438, G0439

WITHOUT

Telehealth Modifier: GQ, GT, 95, POS 02

Population 3:

Patients aged >= 18 years on date of encounter

AND

At least two patient encounters during the performance period (CPT): 90791, 90792, 90832, 90834, 90837, 90845, 92002, 92004, 92012, 92014, 92521, 92522, 92523, 92524, 92540, 92557, 92625, 96150, 96151, 96152, 97165, 97166, 97167, 97168, 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350

WITHOUT

Telehealth Modifier: GQ, GT, 95, POS 02

OR

At least one preventive encounter during the performance period (CPT or HCPCS): 99385, 99386, 99387, 99395, 99396, 99397, 99401, 99402, 99403, 99404, 99411, 99412, 99429, G0438, G0439

WITHOUT

Telehealth Modifier: GQ, GT, 95, POS 02

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

Denominator Exclusions: not applicable. These are identified as denominator exceptions. We have defined the current methodology for distinguishing between denominator exclusions and denominator exceptions in question sp. 18. We did not receive testing information related to these exceptions.

Denominator Exceptions:

Population 1:

Documentation of medical reason(s) for not screening for tobacco use (eg, limited life expectancy, other medical reason)

Population 2:

Documentation of medical reason(s) for not providing tobacco cessation intervention (eg, limited life expectancy, other medical reason)

Population 3:

Documentation of medical reason(s) for not screening for tobacco use OR for not providing tobacco cessation intervention for patients identified as tobacco users (eg, limited life expectancy, other medical reason)

[Response Ends]

sp.18. Provide details needed to calculate the denominator exclusions.

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

[Response Begins]

Time Period for Data Collection: At least once during the measurement period

The PCPI distinguishes between denominator exceptions and denominator exclusions.

Denominator exclusions arise when the clinical action indicated in the numerator is not appropriate for a particular group of patients who otherwise meet the denominator criteria. These are absolute and would be removed from the denominator of a measure in order to determine the eligible population.

Denominator exceptions are used to remove a patient from the denominator when the patient does not receive the action(s) required in the numerator AND that action(s) would not be appropriate due to a patient-specific reason(s). The patient would otherwise meet the denominator criteria. Exceptions are not absolute and are based on provider judgment or individual patient characteristics or preferences. The PCPI methodology includes two categories of exceptions for which a patient may be removed from the denominator of an individual measure: 1) medical OR 2) patient or non-medical reasons. These exception categories are not uniformly relevant across all measures. The denominator exception language may include specific examples of instances that may constitute an exception, which are intended to serve as a guide to providers. 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 providers 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 provider's exceptions data to identify practice patterns and opportunities for quality improvement.

Additional details: This measure includes denominator exceptions.

Population 1:

Report quality data code:

G9904: Documentation of medical reason(s) for not screening for tobacco use (e.g., limited life expectancy, other medical reason)

Population 2:

Report quality data code:

G9907: Documentation of medical reason(s) for not providing tobacco cessation intervention (e.g., limited life expectancy, other medical reason)

Population 3:

Append modifier to CPT Category II code or report quality data code:

4004F-1P: Documentation of medical reason(s) for not screening for tobacco use (e.g., limited life expectancy, other medical reason)

OR

G9909: Documentation of medical reason(s) for not providing tobacco cessation intervention if identified as a tobacco user (e.g., limited life expectancy, other medical reason)

[Response Ends]

sp.19. Provide all information required to stratify the measure results, if necessary.

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

[Response Begins]

Consistent with CMS' Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF, the PCPI encourages collection of race and ethnicity data as well as the results of this measure to be stratified by race, ethnicity, administrative sex, and payer.

[Response Ends]

sp.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

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

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

sp.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.23. Select the appropriate interpretation of the measure score.

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

[Response Begins]

Better quality = Higher score

[Response Ends]

sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

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

[Response Begins]

Calculating the performance rate:

1. Define the initial population. The initial population is identified through a common set of characteristics that define the overall group of patients – or other unit of measurement – targeted for evaluation
2. Define the denominator by identifying the subset of the initial population that meets the denominator criteria. Note: in some cases, the initial population and denominator are identical
3. Determine the numerator by identifying the subset of the denominator that meets the numerator criteria
4. From the patients who did not meet the numerator criteria, determine if the provider has documented whether each patient represents an exception. Subtract from the denominator those patients that meet the conditions for a denominator exception; although the exception cases are removed from the denominator for the measure calculation, the exception rate (i.e., percentage of patients with valid exceptions) should be calculated and reported along with performance rates to highlight variations in care
5. Calculate the performance rate

A patient not meeting the numerator criteria and without a valid and documented exception represents a quality failure

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*
- *The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*
- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

[Response Begins]

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

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

Registry Data

[Response Ends]

sp.31. Identify the specific data source or data collection instrument.

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

[Response Begins]

The reporting entities use claims and registry data for the measure. Each reporting entity may use a different database to pull claims and registry information.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.02. Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

[Response Begins]

No

[Response Ends]

2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]

No additional risk adjustment analysis included

[Response Ends]

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

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

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

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

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

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

AND

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

preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

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

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

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

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

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

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

Definitions

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

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

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

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

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

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75

percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

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

[Response Begins]

Claims

Registry Data

[Response Ends]

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

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

[Response Begins]

Doctors and Clinicians Quality Payment Program PY 2020 Clinician Public Reporting: Measures and Activities from MIPS and QCDR. Found at: https://data.cms.gov/provider-data/dataset/7d6a-e7a6

[Response Ends]

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

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

[Response Begins]

01-01-2020 – 12-31-2020

[Response Ends]

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

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

Please do not select:

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

[Response Begins]

Clinician: Individual

[Response Ends]

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

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

[Response Begins]

The data are from 19,427 physicians and other clinicians (eg nurse practitioners, Physician Assistants). The data were collected from individual providers who opt-in to MIPS.

All 19,427 were included in population 1, those who were screened for tobacco use.

19,234 were included in population 2, the individuals who received tobacco cessation.

15,959 were included in population 3, those who were screened and received an intervention.

[Response Ends]

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

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

[Response Begins]

CMS does not report descriptive data at the patient level.

[Response Ends]

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

[Response Begins]

The same data samples were used for all aspects of testing.

[Response Ends]

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

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

[Response Begins]

CMS does not report patient-level socio-demographic data.

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

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

Choose one or both levels.

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

We utilized the methodology described by John Adams (Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009) to calculate signal-to-noise reliability. This methodology uses the Beta-binomial model to assess how well one can confidently distinguish the performance of one reporting entity from another. Conceptually, the Beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences across reporting entities (plans, physicians, etc.) in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures, such as the flu measure. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across reporting entities.

For the tobacco measure, the provider is the reporting entity (not the plan, as listed below). It is a percentage, bounded by 0 and 100, indicating the proportion of people who were screened for tobacco use that year.

The formula for signal-to-noise reliability is:

$$\text{Signal-to-noise reliability} = \sigma^2_{\text{plan-to-plan}} / (\sigma^2_{\text{plan-to-plan}} + \sigma^2_{\text{error}})$$

Therefore, we need to estimate two variances: 1) variance between plans ($\sigma^2_{\text{plan-to-plan}}$); 2) variance within plans (σ^2_{error}).

1. Variance between plans = $\sigma^2_{\text{plan-to-plan}} = (\alpha\ \beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$

   α and β are two shape parameters of the Beta-Binomial distribution, $\alpha > 0$, $\beta > 0$

1. Variance within plans: $\sigma^2_{\text{error}} = \hat{p}\ (1- \hat{p})/n$

   $\hat{p}$ = observed rate for the plan

   n = plan-specific denominator for the observed rate (most often the number of eligible plan members)

Using Adams' (2009) methodology, we estimated the reliability for each reporting entity, then averaged these reliability estimates across all reporting entities to produce a point estimate of signal-to-noise reliability. We label this point estimate "mean signal-to-noise reliability". The mean signal-to-noise reliability measures how well, on average, the measure can differentiate between reporting entity performance on the measure.

Along with the point estimate of mean signal-to-noise reliability, we are also providing the distribution of the plan-level (and provider-level) signal-to-noise reliability estimates. Each reporting unit's reliability estimate is a ratio of signal to noise, as described above [$\sigma^2_{\text{plan-to-plan}}$ / ($\sigma^2_{\text{plan-to-plan}}$ + σ^2_{error})]. Variability between reporting units ($\sigma^2_{\text{plan-to-plan}}$) is the same for each unit, while the specific reporting unit error (σ^2_{error}) varies. Reliability for each reporting unit is an ordinal measure of how well one can determine where that entity lies in the distribution across reporting units, with higher estimates indicating better reliability.

[Response Ends]

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

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

[Response Begins]

Population 1: Those Screened for Use

We estimated the reliability for each clinician for 2020 performance year reporting. The mean reliability is 0.994.

Reliability Distribution

N	mean	min	10th	25th	50th	75th	90th	max	sdev
19427	0.994	0.888	0.986	0.995	0.998	0.999	1	1	0.012

Population 2: Received an Intervention

We estimated the reliability for each clinician for 2020 performance year reporting. The mean reliability is 0.992.

Reliability Distribution

N	mean	min	10th	25th	50th	75th	90th	max	sdev
19234	0.992	0.887	0.980	0.993	0.997	0.998	0.999	1	0.014

Population 3: Screened and Received an Intervention

We estimated the reliability for each clinician for 2020 performance year reporting. The mean reliability is 0.994.

Reliability Distribution

N	mean	min	10th	25th	50th	75th	90th	max	sdev
15959	0.975	0.896	0.940	0.963	0.981	0.994	1	1	0.023

[Response Ends]

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

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

[Response Begins]

Results indicate very good reliability for all three rates.

[Response Ends]

2b. Validity

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

[Response Begins]

Accountable Entity Level (e.g. hospitals, clinicians)

[Response Ends]

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

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

[Response Begins]

NCQA performed Pearson correlation for construct validity to determine whether the tobacco measure results correlate with another behavioral health screening measure: *Preventive Care and Screening: Unhealthy Alcohol Use: Screening & Brief Counseling*. This test estimates the strength of linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a strong positive linear association: an increase in values of one variable is associated with increase in value of another variable. A value of 0 indicates no linear association. A value of -1 indicates a strong negative relationship in which an increase in values of the first variable is associated with a decrease in values of the second variable. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

The alcohol measure has one rate assessing whether patients were screened for unhealthy alcohol use and received brief counseling if identified as an unhealthy alcohol user. The tobacco has three rates: 1) screened for tobacco use, 2) who received tobacco cessation intervention, and 3) screened and received tobacco cessation intervention. Each tobacco rate was assessed against the alcohol measure rate separately.

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

Population 1: Screened for Tobacco Use

The **Screened for Tobacco Use** rate is positively and moderately associated with the *Preventive Care and Screening: Unhealthy Alcohol Use: Screening & Brief Counseling* measure. The correlation coefficient with the alcohol measure is 0.461 ($p<0.001$).

Population 2: Received Tobacco Cessation Intervention

The **Tobacco Cessation Intervention** rate is positively and moderately associated with the Preventive Care and Screening: Unhealthy Alcohol Use: Screening & Brief Counseling measure. The correlation coefficient with the alcohol measure is 0.371 ($p<0.001$).

Population 3: Screened for Use and Received an Intervention

The **Screened for Use and Received an Intervention** rate is positively and moderately associated with the Preventive Care and Screening: Unhealthy Alcohol Use: Screening & Brief Counseling measure. The correlation coefficient with the alcohol measure is 0.434 ($p<0.001$).

[Response Ends]

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

[Response Begins]

The tobacco measure performance is moderately associated with the *Preventive Care and Screening: Unhealthy Alcohol Use: Screening & Brief Counseling* measure as indicated by the Pearson correlation tests. These findings suggest that clinicians who perform well on one measure will likely perform well on the other which is expected given both measures are assessing preventive behavioral health care.

[Response Ends]

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

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

[Response Begins]

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure.

To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected reporting units from each group (below 25^{th} and above 75^{th} percentiles). The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each reporting unit. The test statistic is then compared against a normal distribution. If the p-value of the test statistic is less than .05, then the two reporting units' performance are significantly different from each other.

[Response Ends]

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

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

[Response Begins]

Population 1: Screened for Tobacco Use:

NCQA calculated the distribution of clinician-level performance for the **Screened for Tobacco Use** rate for the tobacco measure. There is a 49-point gap in performance between the 25th and 75th percentiles. The difference in performance between reporting units in these percentiles is statistically significant.

Validity (t-test) Output

p25	p75	Denominator Bottom Q	Denominator TopQ	Rate LowQ	Rate TopQ	z	p_value_interpret
49	98	1146	770	23	199	58.74	p < 0.001

Population 2: Received Tobacco Cessation Intervention:

NCQA calculated the distribution of clinician-level performance for the **Received Tobacco Cessation Intervention** rate for the tobacco measure. There is a 47-point gap in performance between the 25th and 75th percentiles. The difference in performance between reporting units in these percentiles is statistically significant.

Validity (t-test) Output

p25	p75	Denominator Bottom Q	Denominator TopQ	Rate LowQ	Rate TopQ	z	p_value_interpret
44	91	63	837	16	95	16.88	p < 0.001

Population 3: Screened for Use and Received an Intervention:

NCQA calculated the distribution of clinician-level performance for the Screened for Use and Received an Intervention rate for the tobacco measure. There is a 61-point gap in performance between the 25th and 75th percentiles. The difference in performance between reporting units in these percentiles is statistically significant.

Validity (t-test) Output

p25	p75	Denominator Bottom Q	Denominator TopQ	Rate LowQ	Rate TopQ	z	p_value_interpret
24	85	64	32	20	91	9.98	p < 0.001

[Response Ends]

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

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

[Response Begins]

The difference in performance between reporting units is statistically significant for all three rates.

[Response Ends]

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

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

[Response Begins]

We do not have this information because CMS does not collect this information on the extent and distribution of missing data.

[Response Ends]

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

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

[Response Begins]

We do not have this information because CMS does not collect this information on the extent and distribution of missing data.

[Response Ends]

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

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

[Response Begins]

We do not have this information because CMS does not collect this information on the extent and distribution of missing data.

[Response Ends]

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

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

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

N/A or no exclusions

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

No risk adjustment or stratification

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

Criterion 3. Feasibility

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

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

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

Some data elements are in defined fields in electronic sources

[Response Ends]

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

[Response Begins]

Although the claims data is captured electronically with encounter codes for the denominator and CPT II codes for the numerator, registry implementation may vary.

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

There is an existing eCQM measure. The eCQM number for this measure is CMS138.

[Response Ends]

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

[Response Begins]

We have not identified any areas of concern or made any modifications as a result of testing and operational use 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.

[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]

This Physician Performance Measure (Measure) and related data specifications are owned and copyrighted by the National Committee for Quality Assurance (NCQA). NCQA is not responsible for any use of the Measure. The Measure is

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

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

NCQA makes no representations, warranties, or endorsement about the quality of any organization or physician that uses or reports performance measures and NCQA has no liability to anyone who relies on such measures or specifications. The Measure can be reproduced and distributed, without modification, for noncommercial purposes (e.g., use by healthcare providers in connection with their practices) without obtaining approval from NCQA. Commercial use is defined as the sale, licensing, 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. All commercial uses or requests for modification must be approved by NCQA and are subject to a license at the discretion of NCQA. The PCPI's and AMA's significant past efforts and contributions to the development and updating of the measure are acknowledged.

Limited proprietary coding is contained in the Measure specifications for user convenience. Users of proprietary code sets should obtain all necessary licenses from the owners of the code sets. NCQA disclaims all liability for use or accuracy of any third party codes contained in the specifications.

[Response Ends]

Criterion 4: Use and Usability

4a. Use

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

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

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

4a.01. Check all current uses. For each current use checked, please provide:

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

[Response Begins]

Public Reporting

[Public Reporting Please Explain]

Quality Payment Program (QPP)—

- Sponsor: Centers for Medicare & Medicaid Services (CMS)
- URL: MIPS CQM: https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measures/2022_Measure_226_MIPSCQM.pdf

  eCQM: https://ecqi.healthit.gov/ecqm/ec/2021/cms138v9
- Purpose: To assess tobacco screening and intervention rates
- Geographic area: National
- Level of measurement and setting: Level of measurement: Clinician: Group/Practice, Clinician: Individual; Setting: Outpatient

Payment Program

[Payment Program Please Explain]

Quality Payment Program (QPP)—

- Sponsor: Centers for Medicare & Medicaid Services (CMS)
- URL: MIPS CQM: https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measures/2022_Measure_226_MIPSCQM.pdf

  eCQM: https://ecqi.healthit.gov/ecqm/ec/2021/cms138v9
- Purpose: To assess tobacco screening and intervention rates
- Geographic area: National
- Level of measurement and setting: Level of measurement: Clinician: Group/Practice, Clinician: Individual; Setting: Outpatient

Other (specify)

[Other (specify) Please Explain]

Million Hearts÷

- Sponsor: U.S. Department of Health and Human Services
- URL: https://millionhearts.hhs.gov/files/MH_CQM.pdf
- Purpose: Aligns existing efforts, as well as creates new programs, to improve health across communities and help Americans live longer, more productive lives
- Geographic Area: National

HRSA Uniform Data System

- Sponsor: HRSA
- URL: https://data.hrsa.gov/tools/data-reporting/program-data/national/table?tableName=6B&year=2021
- Purpose: Provide a standardized reporting system for core set of information
- Geographic Area: National
- Level of measurement: Health Centers that receive federal funding as part of HRSA's Health Center Program

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Measure Currently in Use

[Response Ends]

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

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

[Response Begins]

Not applicable

[Response Ends]

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

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

[Response Begins]

Not applicable

[Response Ends]

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

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

[Response Begins]

CMS publishes measure performance results, and scores on its Physician Compare website. Performance year 2020 MIPS scores are publicly available and identifiable by clinician and group. Consumers will be able to see their clinicians rated against national peers on a scale of 0 to 100.

[Response Ends]

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

[Response Begins]

CMS publishes results annually on its Physician Compare website during the year after the performance year.

[Response Ends]

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

[Response Begins]

CMS does not provide information regarding feedback from measured entities on measure performance and implementation.

[Response Ends]

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

[Response Begins]

CMS does not provide information regarding feedback from measured entities on measure performance and implementation.

[Response Ends]

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

[Response Begins]

CMS does not provide information regarding feedback from measured entities on measure performance and implementation.

[Response Ends]

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

[Response Begins]

CMS does not provide information regarding feedback from measured entities on measure performance and implementation.

[Response Ends]

4b. Usability

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

[Response Begins]

Due to limited availability of QPP data, we are not able to demonstrate trends in performance results.

[Response Ends]

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

[Response Begins]

We are not aware of any unintended consequences related to this measure.

[Response Ends]

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

[Response Begins]

We are not yet aware of any unexpected benefits related to this measure.

[Response Ends]

Criterion 5: Related and Competing Measures

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

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

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

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

[Response Begins]

Medical Assistance with Smoking and Tobacco Use Cessation (MSC)

- Steward: NCQA
- Use: HEDIS

[Response Ends]

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

[Response Begins]

No

[Response Ends]

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

[Response Begins]

Related measures have different target populations and/or levels of measurement. Measure 0028 focuses on routine tobacco screening for all adults and tobacco cessation interventions for those who use tobacco products and is intended to assess clinician performance towards these objectives. The cessation intervention required by the measure includes brief counseling and/or pharmacotherapy. Measure 0026 is focused on a target population of infants, children and adolescents. Measure 0027 is a patient survey measure assessing health plan performance and includes one additional component of the cessation intervention beyond this measure (ie, discussion of methods or strategies other than medication). Measures 1651, 1654 and 1656 assess hospital performance. Measure 2803 is focused on assessing clinician performance among adolescents. Measure 2600 represents an adaptation of this measure and is limited to a subset of the population of patients with serious mental illness. Finally, MSC, also from NCQA, only assesses individuals who report smoking by capturing what percentage receive advice from a provider on quitting.

[Response Ends]

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

Provide analyses when possible.

[Response Begins]

No competing measures.

[Response Ends]