NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 2803

Corresponding Measures:

De.2. Measure Title: Tobacco Use and Help with Quitting Among Adolescents

Co.1.1. Measure Steward: National Committee for Quality Assurance

De.3. Brief Description of Measure: The percentage of adolescents 12 to 20 years of age with a primary care visit during the measurement year for whom tobacco use status was documented and received help with quitting if identified as a tobacco user.

1b.1. Developer Rationale: The Tobacco Use and Help with Quitting Among Adolescents measure addresses an issue of significant importance. Tobacco use can have both immediate and long-term serious health consequences, yet data show that, despite some successes, many adolescents continue to begin or use tobacco products. In fact, there was a 38.3% increase in tobacco product use among high school students from 2017 to 2018 according to analysis of the National Youth Tobacco Surveys (Gentzke et al., 2019). The 2017 Youth Risk Behavior Survey of over 14,000 adolescents found that 28.9% had ever tried smoking a cigarette, with 8.8% current smokers, and 42.2% had ever tried an electronic tobacco product, with 13.2% current users (Kann et al., 2018).

According to the Surgeon General, nearly all tobacco use originates in adolescence, with 87% of smokers initiating before the age of 18 (HHS, 2014). The 2017 Youth Risk Behavior Survey found that almost 10% of adolescents had tried cigarette smoking before the age of 13 (Kann et al., 2018). Ceasing tobacco use early in life can reduce risk of premature death to a level comparable to life-time non-smokers (HHS, 2010). Thus, prevention and early intervention in youth are critical to protecting health outcomes throughout the life course.

Research has shown that health care providers play an important role in preventing tobacco use and promoting cessation. This measure encourages consistent documentation of tobacco use status among adolescents and appropriate follow-up for those who are users.

References:

Department of Health and Human Services (HHS). (2010). How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK53017/pdf/Bookshelf_NBK53017.pdf.

Department of Health and Human Services (HHS). (2014). The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General. Atlanta GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK179276/.

Gentzke AS, Creamer M, Cullen KA, et al. (2019). Vital Signs: Tobacco Product Use Among Middle and High School Students — United States, 2011–2018. MMWR Morb Mortal Wkly Rep, 68(6),157–164. Retrieved from https://www.cdc.gov/mmwr/volumes/68/wr/mm6806e1.htm.

Kann L, McManus T, Harris WA, et al. (2018). Youth Risk Behavior Surveillance — United States, 2017. MMWR Surveill Summ, 67(No. SS-8):1–114. Retrieved from https://www.cdc.gov/mmwr/volumes/67/ss/ss6708a1.htm.

S.4. Numerator Statement: Patients who were screened for tobacco use at least once within 18 months (during the measurement period or the six months prior to the measurement period) AND who received tobacco cessation counseling intervention if identified as a tobacco user.

S.6. Denominator Statement: All patients aged 12-20 years with a visit during the measurement period

S.8. Denominator Exclusions: N/A

De.1. Measure Type: Process

S.17. Data Source: Claims, Electronic Health Records

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

IF Endorsement Maintenance – Original Endorsement Date: May 04, 2016 **Most Recent Endorsement Date:** May 04, 2016

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

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

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

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Summary of prior review in 2016

- This measure addresses standardized documentation of tobacco use status among individuals ages 12-20 years adolescents as well as appropriate follow-up for those who are users.
- The evidence from the 2016 review was graded as a B, meaning the USPSTF recommends the service and there is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.
- The evidence for this process measure is based on guidelines/recommendations of two bodies: the U.S. Preventive Services Health Task Force (USPSTF) and the American Academy of Pediatrics (AAP). The developer focuses on the USPSTF recommendations because it is a systematic review; the AAP recommendations are provided within an AAP policy statement.
    - USPSTF Overall Recommendation: The USPSTF recommends that primary care clinicians provide interventions, including education or brief counseling, to prevent initiation of tobacco use among school-aged children and adolescents. Grade B: 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.
    - USPSTF Recommendation 1: Clinicians should ask pediatric and adolescent patients about tobacco use and provide a strong message regarding the importance of totally abstaining from tobacco use. Grade C Strength of Evidence—"reserved for important clinical situations in which the Panel achieved consensus on the recommendation in the absence of relevant randomized controlled trials."
    - USPSTF Recommendation 2: Counseling has been shown to be effective in treatment of adolescent smokers. Therefore, adolescent smokers should be provided with counseling interventions to aid them in quitting smoking. Grade B Strength of Evidence—"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."
    - The systematic review assessed the Quantity, Quality, and Consistency of the literature: 19 trials, of which 4 were rated "good" and 15 were rated "fair".
    - The USPSTF found no evidence on the harms of behavioral interventions to prevent tobacco use and concluded the magnitude of potential harms is probably small to none.
    - Overall, the USPSTF concluded with moderate certainty that primary-care interventions to prevent tobacco use in school-aged children and adolescents have a moderate net benefit.

Changes to evidence from last review

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

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

Updates:

- In the 2020 submission, 7 new trials were identified, 5 of which were rated good quality and 2 of which were rated fair quality.
- Based on the systematic review, the draft recommendation update remains consistent in its recommendation that primary care clinicians provide interventions to prevent initiation of tobacco use among adolescents (B rating).

- Does the Committee agree that the evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review?
- As a documentation only measure, what is the relationship of this measure to patient outcomes? How strong is the evidence for this relationship?

Guidance from the Evidence Algorithm

Box 1: Measure is not assessing a health outcome → Box 3: Evidence is based on a systematic review grading the body of clinical evidence → Box 4: QQC provided → Box 5a.: Evidence is high-quality (From Algorithm 1, NQF Measure Evaluation Criteria Sept 2019, pg. 15)

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The 2017, MIPS performance data suggest that there is variation in performance on this measure and continued opportunity to promote individual clinician and group performance improvement.
- Individual Clinician Performance
    - Mean 91%
    - SD 14%
- Group Clinician Performance
    - Mean 89%
    - SD 17%
- Data from the 2015 National Youth Tobacco Survey (NYTS) showed that only 29% of adolescent respondents reported being asked about tobacco use by a health care provider, with only 22% given advice about not using tobacco products (CDC, 2015).
- From 2017 to 2018, use of e-cigarettes increased by 77.8% among high school students and 48.5% among middle schoolers (Gentzke et al., 2019).

Disparities

- From the 2015 submission, developer provided an analysis of disparities data based on their available sample:
    - Non-Hispanic white and African American adolescents had similar rates of tobacco use/help documented (65.9 and 66.8%, respectively).
    - Hispanic/Latino and Asian/Native American/Pacific Islander adolescents had lower rates (38.1 and 25.0%, respectively).
    - Those of other/multiple races had a rate of 54.6%.
- By insurance status, those with Medicaid had a rate of 60.6%, commercial had a rate of 82.0%, and those who self-paid/had other insurance had a rate of 39.2%.
- An analysis of the 2013 National Survey on Drug Use and Health found that adolescents were more likely to be screened for tobacco use by a clinician if they were white, female, privately insured and in

older adolescence (Collins et al., 2017). The same analysis concluded that Hispanic adolescents were significantly less likely to receive cessation counseling than their peers (Collins et al., 2017).

- From the 2005 National Health Interview Survey
    - White smokers (85%) were slightly more likely to be asked about tobacco use than black (77%) or Hispanic (72%) smokers
    - Minority groups were less likely to be advised to quit (63% in whites, 55% in black and 48% in Hispanics) (Cokkinides et al., 2008).
- Persons (across ages) whose household incomes were below or near the federal poverty level had substantially higher prevalence of smoking, compared with persons whose household incomes were above the federal poverty level. Yet people who have a low socio-economic status are less likely to have adequate access to primary care providers and information about the harms of tobacco use (Fiore et al., 2008).

Questions for the Committee:

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

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

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

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

- Evidence is robust and has been updated.
- Sufficient evidence to support the measure.
- The measure addresses standardized documentation of tobacco use status among individuals 12-20 as well as appropriate follow up for those who are users. The evidence from the 2016 review was a Grade B USPSTF recommends the service and net benefit is moderate. Two guidelines were used: USPSTF and AAP. There was a systematic review of the former but for the latter the recommendations were in a policy statement. For the 2020 submission, there were 7 new trials of which 5 were rated good and 2 were rated fair. The draft recommendation update remains consistent.
- Evidence is strong; updated literature review was provided.
- The USPSTF came out in April 2020 with a Grade B recommendation to use education and brief counseling to prevent usage of tobacco but also stated that the evidence was insufficient to recommend PCP based interventions to cease tobacco use. This SUBSTANTIALLY changes the picture.
- There is evidence to support the measure.
- No concerns; increase in tobacco use noted since most recent endorsement in 2016.
- Generally yes.
- There is sufficient evidence and importance to measure.

- Evidence appears strong. The submission notes that 7 new trials were identified, 5 of which were rated good quality and 2 of which were rated fair quality.
- Process measure.
- The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of primary care–feasible interventions for the cessation of tobacco use among school-aged children and adolescents. This conclusion was reached due to the limited number of high-powered studies looking at behavioral interventions to promote cessation in children and adolescents in the primary care setting.

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

- Not much gap, except disparities. The usual question of are we topped out--we need a better way to focus on disparities.
- Performance gap is demonstrated.
- The 2017 MIPS performance data suggests there is variation in performance in the measure and continual opportunity to promote individual clinician and group performance improvement . Individual clinician mean was 91% and group clinician mean was 89%. As for disparities, in the 2015 submission, non-Hispanic whites and African American adolescents had similar rates of use/help whereas Hispanic and Asian/Pacific Islanders has lower rates. There were also disparities between Medicaid and commercial insurance recipients. There is also data to show that non-whites are less likely to be advised to quit tobacco use.
- Data demonstrates a performance gap.
- It is clear that with the surge in vaping, performance has worsened. It is clear that there are significant economic and racial disparities. There seem to be significant differences in how different individual practitioners and systems of care perform. Given that counseling has been deemed to not be sufficiently evidenced based to warrant USTPF recommendation for smoking cessation (in contrast to preventing use), the logic supporting this measure is no longer valid at this time.
- There does seem to be a gap and some disparities payor and race/ethnicity we reported.
- Yes.
- There is a performance gap.
- The submission includes significant data on performance gap. There appear to be significant racial and ethnic disparities in screening.
- A performance gap exists; disparities exist- white teens provided more brief intervention than black youth.
- Large performance gap. Especially for most at risk.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

2c. For composite measures: empirical analysis support composite approach

Reliability

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

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

Validity

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

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

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

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

Evaluators: NQF Staff

NQF Staff Review

Reliability

- Developer performed reliability testing using a signal-to-noise analysis with the beta-binomial methodology. This is a common approach used for reliability testing at the score level for measures of pass/fail events.
- While the developer has only reported a single summary statistic, developer has met with NQF staff to request the opportunity to share a new approach to include confidence intervals around score level reliability point estimates prior to the Committee meeting.
- Developer was referred to NQF submission criteria: "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." (NQF Measure Evaluation Criteria, pg. 18)
- Developer reports that "Signal-to-noise reliability at the group/practice level is 0.996."

Validity

- Pearson correlation coefficient determined for this measure and *Help Quitting Among Adolescents and Unhealthy Alcohol Use: Screening and Brief Counseling.*
- Face validity was also confirmed as part of 2015 submission.
- Pearson correlation coefficient at the Group/Practice level was 0.37 ($p<0.0001$). This is considered moderate.

Questions for the Committee regarding reliability:

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

- The developer provides a simplistic result that does not include an identifiable interpretation of the point estimate provided, a representation of the spread of reliability results, or other summary statistics. Does the Committee agree with the NQF staff evaluation of the reliability testing for the measure?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- Do you agree with the staff assessment of the validity analyses for the measure?

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

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

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- Moderate reliability--no concerns
- Data elements clearly defined.
- The developer used the signal to noise analysis with a beta binomial methodology but reported only a single summary statistic 0.996 at the group practice level. There was a recent submission demonstrating the reliability testing. There was a subgroup of 75 adolescents and inter rated reliability was demonstrated. It appears that this measure can be consistently implemented.
- I could not find the definitions of what is counted as a tobacco product (e-cigarettes?).
- The reliability specifications are clearly defined.
- No issues.
- I'd like to have discussion on defining counseling and capturing documentation from EHR.
- Specifications clearly defined.
- No concerns.
- The staff recommendation is for "moderate." However, it is unclear from the submission if the USPSTF screening tool distinguishes between e-cigarettes and combustable tobacco.
- No concerns--the measure can be consistently implemented.
- Measure specifications- did the provider use an evidence-based screening tool? Also what cessation interventions were utilized. We should learn which interventions are most effective.

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

- SNR. Usual limitations. Fine.
- No concerns.
- No.
- No concerns re: the testing.
- No.
- No.
- Based only on SNR.

- No.
- No.
- No concerns.
- The developer has only reported a single summary statistic.

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

- Fine.
- No concerns.
- No, there were two methods used: face validity and construct using a Pearson correlation coefficient determined for the measure Help Quitting Among Adolescents and Unhealthy Alcohol Use: Screening and Brief Counseling. The coefficient for group practice was 0.37 which is moderate. Face validity used a multi stakeholder advisory panel which was confirmed as part of the 2015 submission.
- No.
- Validity seems OK.
- No.
- Based only on face validity and correlation with similar measure.
- No.
- Yes - whether the measure is capturing the rapid growth in use of e-cigarettes among adolescents.
- No concerns.
- No.

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

- Ok.
- No concerns.
- Acceptable results, no risk adjustment.
- Concerned as to whether e-cigs are included; developer cites evidence supporting use of screening and counseling for this in the evidence section.
- Given the 2020 USPTF determination that PCP interventions to help school-aged children/adolescents cease tobacco use is NOT sufficiently evidenced based; the overall validity of the measure is lost.
- No threats.
- Does not adjust for social determinants.
- No Concerns.
- I could not find any exclusions in the submission. There appears to be sufficient risk adjustment.
- Analysis indicate acceptable results.
- Unclear about exclusions.

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

- Pretty straight forward about yes/no--but can promote check-boxing. How robust of an intervention is clinically significant.
- No concerns.
- No threats.
- No concerns.
- The developer has only reported a single summary statistic, which seems to lessen the reliability.
- No threats.
- No.
- No.
- No.
- Meaningful differences exist.
- No issues.

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.

- This measure information is captured in electronic health records and is abstracted for reporting.
- Some aspects are reported in structured fields and some can be found in the narrative or in notes.
- There have been no issues reported in the ability of clinicians to collect this data.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- Some elements such as counseling and advice probably require EHR customization.
- Data is available electronically.
- No concerns as the measure information is captured in EHR and is abstracted for reporting. Some aspects are reported in structured fields and some can be found in the narrative or notes. There has been no issue reported in the ability of clinicians to collect the data.

- No concerns.
- It's feasible.
- I did not find problems with any of the data elements.
- See above - discussion surrounding defining counseling and capturing in HER.
- Feasible.
- No concerns.
- I see no problems with feasibility.
- No concerns. Measure is feasible.
- Are the required data elements routinely generated and used during care delivery?

Criterion 4: Usability and Use

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

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

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

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure Publicly reported? ☒ **Yes** ☐ **No**

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

Accountability program details

- Public Reporting
  - CMS Merit-Based Incentive Payment System (data reported publicly via Physician Compare)
- Payment Program
  - CMS Merit-Based Incentive Payment System

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

Developer states that:

- When clinicians and groups report their performance on this measure for MIPS, CMS provides them with feedback reports to inform performance improvement efforts.
- All individual and group performance data is reported publicly through Physician Compare, and annual benchmarks are publicly available to enable clinicians to understand how their performance compares to national benchmarks.

Additional Feedback: None

Questions for the Committee:

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

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

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

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

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

Improvement results

- No improvement results available as measure was tested using 2017 MIPS data, the most recent year available to date.

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

Unexpected findings (positive or negative) during implementation [unexpected findings]

- No unexpected consequences identified. Unexpected benefits may include opening the door to provider-patient conversations about other risky behaviors.

Potential harms

- No potential harms identified.

Additional Feedback: None

Questions for the Committee:

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

Preliminary rating for Usability ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

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

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

- In use and accepted.
- No concerns.
- There is an accountability program involving MIPs. CMS provides feedback reports to inform performance improvement efforts. All individual and group performance data is reported publicly and annual benchmarks are publicly available to enable clinicians to understand how their performance

compares to national benchmarks. No results are available, however, for improvement as the measure was tested using 2017 MIPS data.
- Yes, physicians are given feedback on their MIPS submissions.
- Historical use has been adequate.
- Results have been shared.
- Influenced by documentation effect.
- Yes.
- The submission indicates that this has been posted since 2015 and incorporates feedback on measuring growth in the use e-cigarettes.
- Reported in EHRs. No concerns about use.
- Yes.

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

- Usable and no evidence of unintended consequences.
- No unintended consequences.
- No unexpected consequences and unexpected benefits may include opening the door to provider-patient conversations about other risky behaviors.
- No unexpected consequences.
- This measure is no longer evidence based.
- I believe this would be useful.
- Agree with importance of measure. Would like more data about which counseling and treatment approaches show the most efficacy and improve outcomes.
- Benefits> harm if data interpreted in light of limitations.
- No concerns.
- This measure has high usability. Benefits far exceed any possible harm.
- Performance results can definitely be used to improve care.
- Benefits outweigh harm.

Criterion 5: Related and Competing Measures

Related or competing measures

Measure developer identified the following measure as related: 0028 Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention

Additionally, NQF staff have identified the following measures:

- 0027: Medical Assistance With Smoking and Tobacco Use Cessation
- 2600: Tobacco Use Screening and Follow-up for People with Serious Mental Illness or Alcohol or Other Drug Dependence

Harmonization

These measures have not been harmonized as they address different target populations: 0027, 0028 and 2600 addresses adults 18 and older. The overlap in population between the two measures is minimal and is not expected to create a burden in reporting.

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

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

- Not really--usual question of whether subgroups (e.g., those with mental illness) warrant a separate measure or simply should be reported as a daughter measure.
- Related measures, not different and not competing.
- Related measures 0028. There are also 2 other measures, 0027 and 2600 but these are focused on adults over 18.
- The other measures related to smoking cessation address adult populations.
- 0027: Medical Assistance With Smoking and Tobacco Use Cessation • 2600: Tobacco Use Screening and Follow-up for People with Serious Mental Illness or Alcohol or Other Drug Dependence are related measures but do not really overlap enough to be easily harmonized.
- There are two other measure but they are targeted to other populations.
- No.
- No.
- Yes, there were 2 competing measures listed, with no harmonization.
- Other related measures do not target adolescents as this one does.
- Measure should be harmonized with the other measures that are for 18 year olds and over.

Public and Member Comments

No Comments and Member Support/Non-Support Submitted as of: 06/05/2020

NQF Staff Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2803

Measure Title: Tobacco Use and Help with Quitting Among Adolescents

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**
☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☒ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☒ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☐ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: Specifications, items S.1-S.22

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

2. **Briefly summarize any concerns about the measure specifications.**
   - None identified

RELIABILITY: TESTING

Submission document: Testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2
   - Developer performed reliability testing using a signal-to-noise analysis with the beta-binomial methodology.
   - This is a common approach used for reliability testing at the score level for measures of pass/fail events.

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

   Submission document: Testing attachment, section 2a2.3
   - While the developer has only reported a single summary statistic, developer has met with NQF staff to request the opportunity to share a new approach to include confidence intervals around score level reliability point estimates prior to the Committee meeting.
   - Developer was referred to NQF submission criteria: "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." (NQF Measure Evaluation Criteria, pg. 18)
   - Developer reports that "Signal-to-noise reliability at the group/practice level is 0.996."

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - There are no exclusions. No concerns

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

    Submission document: Testing attachment, section 2b4.

    - Inter-quartile range calculated to determine difference between high and low-performing practices. IQR was statistically significant and represented a difference of 95 patients, on average.

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

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

    - Missing data was not assessed.

    Submission document: Testing attachment, section 2b6.

16. **Risk Adjustment**

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

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

    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

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

    16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☒ No

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

16d. **Risk adjustment summary:**

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

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

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

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

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

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

This measure is not risk adjusted.

VALIDITY: TESTING

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

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

☒ Face validity

☒ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

- Pearson correlation coefficient determined for this measure and *Help Quitting Among Adolescents and Unhealthy Alcohol Use: Screening and Brief Counseling.*
- Face validity was also confirmed as part of 2015 submission.

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

Submission document: Testing attachment, section 2b2.3

- Pearson correlation coefficient at the Group/Practice level was 0.37 ($p<0.0001$). This is considered moderate.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

NOTE that data element validation from the literature is acceptable.

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

☐ **Yes**

☐ **No**

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

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

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

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

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

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

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

    Box 1: Threats to validity empirically assessed →Box 2: Empirical validity testing conducted using measure as specified and appropriate tests → Box 5: Performance measure score testing → Box 6: Correlation of performance measure score tested against related performance measure score → Box 7b. Moderate confidence that performance measure score indicates quality

ADDITIONAL RECOMMENDATIONS

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

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

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

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

nqf_evidence_attachment_7.1-637214238770221818.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** NQF #2803

Measure Title: **Tobacco Use and Help with Quitting Among Adolescents**

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 4/2/2020

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

Outcome

☐ Outcome: Click here to name the health outcome

☐ Patient-reported outcome (PRO): Click here to name the PRO

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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Adolescent Tobacco Cessation

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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

2015 submission:

Adolescent is seen by physician >> Physician assesses the adolescent as a tobacco user or non-tobacco user >> If adolescent is a tobacco user, physician provides assistance with quitting >> Adolescent ceases using tobacco >> Adolescent's risk of developing tobacco-related morbidity/mortality decreases

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

2020 submission:

Results from the 2019 National Youth Tobacco Survey, a nationally representative survey of 22,153 students, indicate that adolescents believe that tobacco use is harmful to their health. When asked, "How much do you think people harm themselves when they smoke cigarettes some days but not every day?" 54.3% responded "A lot of harm" and 35.2% responded "some harm." When asked the same question about e-cigarette use, 31.9% believed e-cigarettes cause a lot of harm, and 38.8 believed they caused at least some harm. The survey also asked, "How strongly do you agree with the statement 'All tobacco products are dangerous'?" 48.7% responded "strongly agree" and 36.6% agreed (CDC, 2019). These results indicate that a substantial proportion of adolescents do believe tobacco use is harmful to health. The survey did not ask specifically about beliefs or perceptions related to screening and counseling.

References:

Centers for Disease Control and Prevention (CDC). (2019). *National Youth Tobacco Survey* [Dataset and Codebook]. Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/data/index.html

Centers for Disease Control and Prevention Office on Smoking and Health. (2019). *2019 National Youth Tobacco Survey: Methodology Report.* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Retrieved from https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/data/index.html

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

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

N/A

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

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses

explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☒ Clinical Practice Guideline recommendation (with evidence review)

☒ US Preventive Services Task Force Recommendation

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

☐ Other

We have provided 2 clinical practice guidelines and 1 USPSTF recommendation to support this measure: The U.S. Preventive Services Task Force (USPSTF) Guideline, which focuses on primary care interventions to prevent tobacco use initiation in children and adolescents, the U.S. Public Health Service, which recommends clinicians discuss the risk of tobacco use and offer counseling on tobacco cessation to active tobacco users and the American Academy of Pediatrics policy statement on tobacco use.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**2015 submission:** **In this section, we focus on the findings of the systematic review of the evidence upon which the U.S. Preventive Services Task Force based its recommendation related to primary care interventions to prevent tobacco use initiation in children and adolescents** **Title: Final Update Summary: Tobacco Use in Children and Adolescents: Primary Care Interventions** **Author: U.S. Preventive Services Task Force** **Date: July 2015** **URL: http://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/tobacco-use-in-children-and-adolescents-primary-care-interventions**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**2015 submission:** The systematic review outlined in this section supports the following USPSTF recommendation: The USPSTF recommends that primary care clinicians provide interventions, including education or brief counseling, to prevent initiation of tobacco use among school-aged children and adolescents. The Evidence-based Practice Center (EPC) that conducted the review assessed the evidence for the efficacy and harms of primary care-relevant interventions that aim to reduce tobacco use among children and adolescents. Though the USPSTF focuses on primary-care based interventions and our measure includes assessment, in the USPSTF guideline, the assessment is assumed and a logical necessary step towards providing interventions.

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**2015 submission:**
	
	Evidence included 19 trials of which four were good-quality and 15 were fair-quality trials.
	
	Randomized Controlled Trials and Cohort Studies Criteria:
	• Initial assembly of comparable groups:
	  ○ For RCTs: adequate randomization, including first concealment and whether potential confounders were distributed equally among groups.
	  ○ For cohort studies: consideration of potential confounders with either restriction or measurement for adjustment in the analysis; consideration of inception cohorts.
	• Maintenance of comparable groups (includes attrition, cross-overs, adherence, contamination).
	• Important differential loss to follow-up or overall high loss to follow-up.
	• Measurements: equal, reliable, and valid (includes masking of outcome assessment).
	• Clear definition of interventions.
	• All important outcomes considered.
	• Analysis: adjustment for potential confounders for cohort studies, or intention to treat analysis for RCTs.
	
	Definition of ratings based on above criteria:
	• Good: Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention to confounders in analysis. In addition, for RCTs, intention to treat analysis is used.
	• Fair: Studies will be graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention to treat analysis is done for RCTs.

Provide all other grades and definitions from the evidence grading system	**2015 submission:** In addition to *Good* and *Fair*, the evidence-based practice center includes a category of *Poor* for RCTs and Cohort Studies Poor: Studies will be graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat analysis is lacking.

Grade assigned to the **recommendation** with definition of the grade	**2015 submission:** Grade B: The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.
Provide all other grades and definitions from the recommendation grading system	**2015 submission:** *Grade A*: The USPSTF recommends this service. There is high certainty that the net benefit is substantial. *Grade C*: The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small *Grade D*: 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. Discourage the use of this service. I statement. *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. http://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions

Body of evidence:
• Quantity – how many studies?
• Quality – what type of studies?
	2015 submission:
	
	The EPC identified 19 trials that examined the efficacy of primary care-relevant interventions in preventing tobacco use initiation, promoting cessation among youth, and/or harms of interventions. While the study designs varied, all were rated as Good or Fair quality by the EPC. Below is a general description, per topic studied, of the trials that were assessed.
	• Combined prevention and cessation interventions
	  ○ 7 trials: 5 randomized controlled trials (RCTs), 2 cluster-randomized trials
	• Prevention interventions
	  ○ 10 trials: 4 of the trials from the assessment for combined treatment were included in addition to 6 trials on behavior-based interventions to prevent tobacco initiation
	  ○ Two of the studies were rated good quality based on their methods (e.g., valid randomization techniques, good intervention fidelity)
	  ○ The remaining were rated as fair quality, as randomization procedures were not reported or uncertain
	  ○ The EPC noted there was also a lack of blinding for outcome assessors but concluded this was unlikely to produce bias in the studies using standardized data collection tools (e.g., computer-assisted telephone interviewing)
	• Cessation interventions (Behavior and Bupropion)

	○ 10 trials: 9 trials examined cessation among baseline smokers. Of these, two were rated good quality based on their methods (e.g., valid randomization techniques and good intervention fidelity) ○ The remaining were rated fair quality due to issues including attrition and concerns with participant compliance • Adverse effects associated with interventions ○ None of the trials of behavior-based interventions explicitly reported on treatment harms, but three medication-specific studies reported on harms ○ All three medication-specific trials included randomization techniques. Fair to good quality: most of the trials included in the evidence review included randomization and had good intervention fidelity.
Estimates of benefit and consistency across studies	**2015 submission:** The USPSTF found adequate evidence that behavioral counseling interventions can reduce the risk of smoking initiation in school-aged children and adolescents.
What harms were identified?	**2015 submission:** The USPSTF found no evidence on the harms of behavioral interventions to prevent tobacco use and concluded the magnitude of potential harms is probably small to none. Overall, the USPSTF concluded with moderate certainty that primary-care interventions to prevent tobacco use in school-aged children and adolescents have a moderate net benefit.

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	**2020 submission:** New evidence identified since previous submission: The USPSTF is in the process of updating their recommendation related to primary care interventions to prevent initiation of tobacco use in adolescents. The updated recommendation is currently in Draft, and the final version has not yet been published following a public comment period. The USPSTF commissioned a systematic review to support this update. 7 new trials were identified, 5 of which were rated good quality and 2 of which were rated fair quality. Based on the systematic review, the draft recommendation update remains consistent in its recommendation that primary care clinicians provide interventions to prevent initiation of tobacco use among adolescents (B rating). The draft adds an additional recommendation of "I" for cessation interventions, stating "The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of primary care–feasible interventions for the cessation of tobacco use among school-aged children and adolescents." This conclusion was reached due to the limited number of high-powered studies looking at behavioral interventions to promote cessation in children and adolescents in the primary care setting. Of note in the updated recommendation is the inclusion of Electronic Nicotine Delivery Systems (ENDS) as a tobacco product, consistent with the updated FDA classification of ENDS as a tobacco product. The new recommendation concludes that all recommendations related to tobacco use products should be applied to ENDS as well as cigarette products. The inclusion of ENDS in the USPSTF's recommendation underscores the importance of continued screening and counseling around tobacco use among adolescents, considering the dramatic increase in ENDS use among this population in recent years. References: U.S. Preventive Services Task Force. (2019). *Draft Recommendation Statement: Prevention and Cessation of Tobacco Use in Children and Adolescents: Primary Care Interventions*. U.S. Preventive Services Task Force. Retrieved from https://www.uspreventiveservicestaskforce.org/Page/Document/draft-recommendation-statement/tobacco-and-nicotine-use-prevention-in-children-and-adolescents-primary-care-interventions Selph, S., Patnode, C., Bailey, S., Pappas, M., Stoner, R., Hart, E., & Chau, R. (2019). *Systematic Review: Primary Care Relevant Interventions for Tobacco and Nicotine Use Prevention and Cessation in Children and Adolescents: A Systematic Review for the U.S. Preventive Services Task Force.* Unpublished

	manuscript. Retrieved from https://www.uspreventiveservicestaskforce.org/Home/GetFile/1/17030/tobacco-use-prevention-children-draft-evidence-review/pdf

Source of Systematic Review:	**2015 submission:**

• **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**In this section, we focus on the findings of the systematic review of the evidence upon which the U.S. Public Health Service based its recommendations.** **Fiore MC, Jaén 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.** **http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/clinicians-providers/guidelines-recommendations/tobacco/clinicians/update/treating_tobacco_use08.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**2015 submission:** *Recommendation 1*: Clinicians should ask pediatric and adolescent patients about tobacco use and provide a strong message regarding the importance of totally abstaining from tobacco use. *Recommendation 2*: Counseling has been shown to be effective in treatment of adolescent smokers. Therefore, adolescent smokers should be provided with counseling interventions to aid them in quitting smoking.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**2015 submission:** *Recommendation 1*: Strength of Evidence C - Reserved for important clinical situations in which the Panel achieved consensus on the recommendation in the absence of relevant randomized controlled trials. Recommendation 2: Strength of Evidence 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.
Provide all other grades and definitions from the evidence grading system	**2015 submission:** Strength of Evidence A: Multiple well-designed randomized clinical trials, directly relevant to the recommendation, yielded a consistent pattern of findings.
Grade assigned to the **recommendation**	**2015 submission:** *Recommendation 1*: Strength of Evidence C - Reserved for important clinical situations in which the Panel achieved consensus on the recommendation in the absence of relevant randomized controlled trials.

with definition of the grade	*Recommendation 2*: Strength of Evidence 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.
Provide all other grades and definitions from the recommendation grading system	**2015 submission:** Strength of Evidence A: Multiple well-designed randomized clinical trials, directly relevant to the recommendation, yielded a consistent pattern of findings.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	7 studies met inclusion criteria. Based on the quality of the evidence, the Public Health Service rated their recommendation "B" for providing treatment to tobacco users, and "C" for screening and providing preventive counseling to adolescents.
Estimates of benefit and consistency across studies	The U.S. Public Health Service reached a similar conclusion as the USPSTF, that behavioral counseling interventions can reduce the risk of smoking initiation in school-aged children and adolescents. The Public Health Service also recommends clinicians provide adolescent tobacco users assistance with quitting, citing research that has shown that a provider's advice to quit can be effective.
What harms were identified?	The Public Health Service found no evidence of harms.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	We are not aware of any major new evidence reviews conducted by the U.S. Public Health Service since this systematic review.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**2015 submission:** **American Academy of Pediatrics. Committee on Environmental Health, Committee on Substance Abuse, Committee on Adolescence, and Committee on Native American Child Health. 2009. Tobacco Use: A Pediatric Disease. Pediatrics 124(5): 1474. [Reaffirmed May 2013]** **2020 submission:**

	This section focuses on the evidence review prepared to inform the American Academy of Pediatrics Clinical Practice Policy "Protecting Children from Tobacco, Nicotine, and Tobacco Smoke." **References** **Farber, H., Groner, J., Walley, S., Nelson, K. and the Section on Tobacco Control. (2015). Protecting Children from Tobacco, Nicotine, and Tobacco Smoke. *Pediatrics.136*(5): e1439-e1467. Retrieved from https://pediatrics.aappublications.org/content/136/5/e1439?ijkey=6654c76314534a3548b728837254fec981a89ef3&keytype2=tf_ipsecsha** **American Academy of Pediatrics (AAP) Section on Tobacco Control. (2015). Clinical Practice Policy to Protect Children from Tobacco, Nicotine, and Tobacco Smoke. *Pediatrics.136*(5): 1008-1017. Retrieved from https://pediatrics.aappublications.org/content/136/5/1008**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**2015 submission:** For patients and their family members: • Counsel children and parents about the harms of tobacco use. • Include tobacco in all discussions of substances of abuse and risky behaviors. Discussion and anticipatory guidance about tobacco use should ideally begin by 5 years of age and emphasize resisting the influence of advertising and rehearsal of peer refusal skills. Be aware of confidentiality issues related to tobacco use and other substance abuse, including testing for nicotine and its metabolites. • Encourage parents to start discussions of tobacco use with their children early in their life and continue to do so throughout childhood and adolescence; these discussions should include delivery of clear messages disapproving of tobacco use. Both parents and children should be counseled that it is not safe to "experiment" with tobacco, because nicotine is so highly addictive and there is no safe way to use tobacco. Tobacco dependence can begin almost as soon as use begins, with some users exhibiting signs of dependence with only occasional or monthly use. As a result, prevention of tobacco use is one of the most important messages you can deliver. For patients or family members who use tobacco • Advise all families to make their homes and cars smoke free, and urge all tobacco users to quit. Provide appropriate advice and counseling to foster tobacco users to quit. Routinely offer help and referral to those who use tobacco— even if the person is not your patient. Be familiar with evidence-based guidelines for treatment of tobacco use and dependence and apply them to patients and their families.14 There is a growing body of literature on the effectiveness of pediatric clinician-provided treatment for parental nicotine addiction that demonstrates a role for pediatricians in this effort. ○ Pharmacotherapy is an effective component of tobacco use-cessation treatment in adults. Encourage tobacco users to include these medications in their quit plan, whenever appropriate. Be familiar with and offer information and instruction on correct use. Many nicotine replacement products are available without a prescription, although prescriptions are required for any nicotine-containing product if the patient is younger than 18 years.

<table>
<tr>
<td></td>
<td>

Pediatricians who choose not to prescribe pharmacotherapies should make referrals to cessation services and recommend that parents discuss pharmacotherapies with their health care providers or purchase over-the-counter products.

Be familiar with tobacco use– cessation services in your community and provide referrals to these programs for your patients and their families. Memorize the national quit line telephone number (1-800-QUIT NOW), prominently post it, and provide it to all tobacco users. Whenever possible, proactively enroll tobacco users in cessation programs, using “fax-back” or similar programs. Such referrals are more effective in connecting the tobacco user to the resource than referrals that require the tobacco user to initiate the contact.
Counsel all parents, including those who smoke, on how to deliver anti-tobacco messages and ways to discuss the addictive nature of nicotine.

When parents or caregivers use tobacco, their children are more likely to experiment with tobacco and to begin to use tobacco regularly. Maintain a high index of suspicion for early onset of tobacco use by these children. It can be a particularly powerful message when the parent or caregiver who uses tobacco advises the child never to start using tobacco.
Help patients and families understand that even casual use of tobacco by children and adolescents, regardless of amount or frequency, is illegal and associated with adverse health consequences.

<u>2020 submission:</u>

Recommended Actions for Pediatricians
1. Inquire about tobacco use and tobacco smoke exposure as part of health supervision visits and visits for diseases that may be caused or exacerbated by tobacco smoke exposure.
Pediatricians need to be aware of the different terminology that families may use for tobacco products. Because many families may not consider electronic nicotine delivery systems as “tobacco,” questions may need to be modified to include “vape” or “vaping” and/or use of electronic cigarettes, hookah sticks, e-hookahs, and/or vape pens.
2. Include tobacco use prevention as part of anticipatory guidance.
4. Offer tobacco dependence treatment and/or referral to adolescents who want to stop smoking.

5. Offer tobacco-dependent individuals quitline referral.
</td>
</tr>
<tr>
<td>Grade assigned to the evidence associated with the recommendation with the definition of the grade</td>
<td>
<table>
<tr><th>Recommendation</th><th>Evidence Quality</th></tr>
<tr><td>1. Inquire about tobacco use and tobacco smoke exposure as part of health supervision visits and visits for diseases that may be caused or exacerbated by tobacco smoke exposure.</td><td>B</td></tr>
<tr><td>2. Include tobacco use prevention as part of anticipatory guidance.</td><td>B</td></tr>
<tr><td>4. Offer tobacco dependence treatment and/or referral to adolescents who want to stop smoking.</td><td>B</td></tr>
<tr><td>5. Offer tobacco-dependent individuals quitline referral.</td><td>A</td></tr>
</table>
Evidence Quality Definitions:
</td>
</tr>
</table>

<table>
<tr><td></td><td>A. RCTs or diagnostic studies with minor limitations; overwhelmingly consistent evidence from observational studies.
B. Observational studies (case-control and cohort design)</td></tr>
<tr><td>Provide all other grades and definitions from the evidence grading system</td><td><u>2020 submission:</u>

Additional evidence Quality Definitions:
C. Expert Opinion, case reports, reasoning from first principles.
X. Exceptional situations in which validating studies cannot be performed and there is a clear preponderance of benefit or harm.</td></tr>
<tr><td>Grade assigned to the recommendation with definition of the grade</td><td><u>2020 submission:</u>

All recommendations cited above were rated “Strong Recommendation.”
<table><tr><th>Statement</th><th>Definition</th><th>Implication</th></tr><tr><td>Strong recom-mendation</td><td>A strong recommendation in favor of a particular action is made when the anticipated benefits of the recommended intervention clearly exceed the harms (as a strong recommendation against an action is made when the anticipated harms clearly exceed the benefits) and the quality of the supporting evidence is excellent. In some clearly identified circumstances, strong recommendations may be made when high-quality evidence is impossible to obtain and the anticipated benefits strongly outweigh the harms</td><td>Clinicians and policy makers should follow a strong recommendation unless a clear and compelling rationale for an alternative approach is present</td></tr></table></td></tr>
<tr><td>Provide all other grades and definitions from the recommendation grading system</td><td><u>2020 submission:</u>

<table><tr><th>Statement</th><th>Definition</th><th>Implication</th></tr><tr><td>Recom-mendation</td><td>A recommendation in favor of a particular action is made when the anticipated benefits exceed the harms, but the quality of evidence is not as strong. Again, in some clearly identified circumstances, recommendations may be made when high-quality evidence is impossible to obtain but the anticipated benefits outweigh the harms</td><td>Clinicians and policy makers would be prudent to follow a recommendation but should remain alert to new information and sensitive to patient preferences</td></tr><tr><td>Option</td><td>Options define courses that may be taken when either the quality of evidence is suspect or carefully performed studies have shown little clear advantage to one approach over another</td><td>Clinicians and policy makers should consider the option in their decision-making, and preference may play a substantial role</td></tr></table></td></tr>
</table>

	No recom-mendation	No recommendation indicates that there is a lack of pertinent published evidence and that the anticipated balance of benefits and harms is presently unclear	Clinicians and policy makers should be alert to new published evidence that clarifies the balance of benefit versus harm

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	**2020 submission:** The AAP policy statement relies on evidence from the USPSTF systematic review and the US Public Health Service systematic review, discussed above. In addition, the policy statement cites findings from the Memphis Health Project, a longitudinal study of over 5000 adolescents, and the National Youth Tobacco Surveys of 2000 and 2011, covering 24,573 and 18,866 adolescents, respectively. The AAP graded the quality of the evidence "B" according to the criteria described above. Based on the AAPs evaluation of the evidence, a Strong Recommendation was assigned to this policy statement, indicating that the evidence quality is considered excellent.
Estimates of benefit and consistency across studies	**2020 submission:** According to the AAP policy *on Classifying Recommendations for Clinical Practice Guidelines,* "A strong recommendation means that the committee believes that the benefits of the recommended approach clearly exceed the harms of that approach (or, in the case of a strong negative recommendation, that the harms clearly exceed the benefits) and that the quality of the evidence supporting this approach is either excellent or impossible to obtain. Clinicians should follow such guidance unless a clear and compelling rationale for acting in a contrary manner is present." All the relevant recommendation statements included in this section are rated "Strong Recommendation," indicating that the AAP found that benefits clearly outweigh harms. References American Academy of Pediatrics (AAP) Steering Committee on Quality Improvement and Management. (2015). Classifying Recommendations for Clinical Practice Guidelines. *Pediatrics. 114*(3): 874-877. Retrieved from https://pediatrics.aappublications.org/content/114/3/874
What harms were identified?	**2020 submission:** The AAP technical report does not identify any harms associated with these recommendation statements.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	**2020 submission:** We are not aware of any major new evidence reviews conducted by the AAP since this technical report/evidence review.

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

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

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

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

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

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

The Tobacco Use and Help with Quitting Among Adolescents measure addresses an issue of significant importance. Tobacco use can have both immediate and long-term serious health consequences, yet data show that, despite some successes, many adolescents continue to begin or use tobacco products. In fact, there was a 38.3% increase in tobacco product use among high school students from 2017 to 2018 according to analysis of the National Youth Tobacco Surveys (Gentzke et al., 2019). The 2017 Youth Risk Behavior Survey of over 14,000 adolescents found that 28.9% had ever tried smoking a cigarette, with 8.8% current smokers, and 42.2% had ever tried an electronic tobacco product, with 13.2% current users (Kann et al., 2018).

According to the Surgeon General, nearly all tobacco use originates in adolescence, with 87% of smokers initiating before the age of 18 (HHS, 2014). The 2017 Youth Risk Behavior Survey found that almost 10% of adolescents had tried cigarette smoking before the age of 13 (Kann et al., 2018). Ceasing tobacco use early in life can reduce risk of premature death to a level comparable to life-time non-smokers (HHS, 2010). Thus, prevention and early intervention in youth are critical to protecting health outcomes throughout the life course.

Research has shown that health care providers play an important role in preventing tobacco use and promoting cessation. This measure encourages consistent documentation of tobacco use status among adolescents and appropriate follow-up for those who are users.

References:

Department of Health and Human Services (HHS). (2010). How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for

Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK53017/pdf/Bookshelf_NBK53017.pdf.

Department of Health and Human Services (HHS). (2014). The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General. Atlanta GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK179276/.

Gentzke AS, Creamer M, Cullen KA, et al. (2019). Vital Signs: Tobacco Product Use Among Middle and High School Students — United States, 2011–2018. MMWR Morb Mortal Wkly Rep, 68(6),157–164. Retrieved from https://www.cdc.gov/mmwr/volumes/68/wr/mm6806e1.htm.

Kann L, McManus T, Harris WA, et al. (2018). Youth Risk Behavior Surveillance — United States, 2017. MMWR Surveill Summ, 67(No. SS-8):1–114. Retrieved from https://www.cdc.gov/mmwr/volumes/67/ss/ss6708a1.htm.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

The following data are extracted from publicly available feedback on individual clinician and group performance against this measure for MIPS performance year 2017 (the only year of data that CMS has made publicly available on Physician Compare). In 2017, 3043 individual clinicians and 512 groups reported this measure.

> In 2017, eligible clinicians (EC) could report quality measures for periods of 90 days to 12 months. Three-quarters reported data for a full 12-month period, and a quarter reported for periods of at least 90 days (CMS, 2018).
>
> | | MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH |
> |---|---|---|---|---|---|---|---|---|---|
> | INDIVIDUAL | 91.13% | 14.38% | 1.00% | 100.00% | | | 77.00% | 89.75% | 96.00% | 99.00% | 100.00% |
> | GROUP | 89.12% | 17.24% | 4.00% | 100.00% | | 70.10% | 87.00% | 96.00% | 99.00% | 100.00% | |
>
> The 2017, MIPS performance data show that there is variation in performance on this measure, and continued opportunity to promote clinician and group performance improvement.

References:

Centers for Medicare and Medicaid Services (CMS). (2018). 2017 Quality Payment Program Reporting Experience. Baltimore, MD: Centers for Medicare and Medicaid Services. Retrieved from: https://www.pcpcc.org/resource/2017-quality-payment-program-reporting-experience

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

Clinical practice guidelines recommend that clinicians provide tobacco use screening, education and counseling to adolescents to prevent initiation of tobacco products and promote cessation among users (Fiore et al. 2008, Moyer et al. 2013, Farber et al. 2015). However, data from the 2015 National Youth Tobacco Survey (NYTS) showed that only 29% of adolescent respondents reported being asked about tobacco use by a health care provider, with only 22% given advice about not using tobacco products (CDC, 2015).

This is concerning, because while adolescent smoking has decreased over time, overall consumption of tobacco products has increased among adolescents. This may be driven in large part by increased usage of alternative delivery systems, like electronic nicotine delivery systems (ENDS) (El-Toukhy, Sabado & Choi, 2018). From 2017 to 2018, use of e-cigarettes increased by 77.8% among high school students and 48.5% among middle schoolers (Gentzke et al., 2019). Compounding this problem is research showing ENDS use is associated

with future cigarette initiation (Soneji et al., 2017). Given the rising use of electronic tobacco products among youth, in 2016 the FDA extended its definition of tobacco products to include ENDS, making them subject to the FDA's tobacco control authority (FDA, 2016). The American Academy of Pediatrics (AAP) policy on tobacco control now recommends pediatricians screen for and provide counseling for both combustible and e-cigarette use (AAP, 2015).

Given the rise in youth tobacco use in recent years, accompanied with relatively low patient reports of clinician screening and counseling, this measure serves an important role in promoting the continued documentation of tobacco use status, including ENDS, and the role of clinicians in promoting abstinence and cessation.

References:

American Academy of Pediatrics Section on Tobacco Control (AAP). (2015). Policy statement: Electronic Nicotine Delivery Systems. Pediatrics.136(5):1018—1026. Retrieved from https://pediatrics.aappublications.org/content/pediatrics/136/5/1018.full.pdf

Centers for Disease Control and Prevention (CDC). (2015). National Youth Tobacco Survey [Dataset and Codebook]. Centers for Disease Control and Prevention. Retrieved from https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/data/index.html

El-Toukhy S, Sabado M, Choi K. (2018). Trends in tobacco product use patterns among U.S. youth, 1999–2014. Nicotine Tob Res. 20(6):690–697. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28582567.

Food and Drug Administration (FDA). (2016). 81 FR 28973 Deeming Tobacco Products to Be Subject to the Federal Food, Drug, and Cosmetic Act, as Amended by the Family Smoking Prevention and Tobacco Control Act; Restrictions on the Sale and Distribution of Tobacco Products and Required Warning Statements for Tobacco Products. Silver Spring, MD: Office of Regulations, Center for Tobacco Products, Food and Drug Administration. Retrieved from https://www.federalregister.gov/documents/2016/05/10/2016-10685/deeming-tobacco-products-to-be-subject-to-the-federal-food-drug-and-cosmetic-act-as-amended-by-the

Farber, H.J., Walley, S.C., Groner, J.A. et al. (2015). Clinical practice policy to protect children from tobacco, nicotine, and tobacco smoke. Pediatrics.136(5): 1008–1017. Retrieved from https://pediatrics.aappublications.org/content/136/5/1008.

Fiore MC, Jaén CR, Baker TB, et al. (2008). Treating Tobacco Use and Dependence: 2008 Update. Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services. Public Health Service. Retrieved from https://www.ahrq.gov/prevention/guidelines/tobacco/index.html.

Gentzke AS, Creamer M, Cullen KA, et al. (2019). Vital Signs: Tobacco Product Use Among Middle and High School Students — United States, 2011–2018. MMWR Morb Mortal Wkly Rep, 68(6),157–164. Retrieved from https://www.cdc.gov/mmwr/volumes/68/wr/mm6806e1.htm.

Moyer VA, et al. (2013). Primary care interventions to prevent tobacco use in children and adolescents: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 159(8):552–557. Retrieved from https://annals.org/aim/fullarticle/1748857/primary-care-interventions-prevent-tobacco-use-children-adolescents-u-s.

Soneji S, Barrington-Trimis JL, Wills TA, et al. (2017). Association between initial use of e-cigarettes and subsequent cigarette smoking among adolescents and young adults: a systematic review and meta-analysis. JAMA Pediatr. 171(8):788–797. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28654986.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(<u>This is required for maintenance of endorsement</u>. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We have some data on disparities from field testing conducted during the measure's development. Across three clinician sites, the percentage of adolescents whose tobacco use was documented and, if smokers, who received help with quitting varied by race/ethnicity and insurance status. In our sample, non-Hispanic white and African American adolescents had similar rates of tobacco use/help documented (65.9 and 66.8%, respectively). Hispanic/Latino and Asian/Native American/Pacific Islander adolescents had lower rates (38.1 and 25.0%, respectively). Those of other/multiple races had a rate of 54.6%.

By insurance status, those with Medicaid had a rate of 60.6%, commercial had a rate of 82.0%, and those who self-paid/had other insurance had a rate of 39.2%.

Data used to conduct testing in 2020 did not enable analysis by population group.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

An analysis of the 2013 National Survey on Drug Use and Health found that adolescents were more likely to be screened for tobacco use by a clinician if they were white, female, privately insured and in older adolescence (Collins et al., 2017). The same analysis concluded that Hispanic adolescents were significantly less likely to receive cessation counseling than their peers (Collins et al., 2017).

From the 2005 National Health Interview Survey, white smokers (85%) were slightly more likely to be asked about tobacco use than black (77%) or Hispanic (72%) smokers (Cokkinides et al., 2008). The survey also indicated that minority groups were less likely to be advised to quit (63% in whites, 55% in black and 48% in Hispanics) (Cokkinides et al., 2008). Persons (across ages) whose household incomes were below or near the federal poverty level had substantially higher prevalence of smoking, compared with persons whose household incomes were above the federal poverty level. Yet people who have a low socio-economic status are less likely to have adequate access to primary care providers and information about the harms of tobacco use (Fiore et al., 2008).

Collins, L., Smiley, S. L., Moore, R. A., Graham, A. L., & Villanti, A. C. (2017). Physician tobacco screening and advice to quit among U.S. adolescents - National Survey on Drug Use and Health, 2013. Tobacco induced diseases, 15, 2. doi:10.1186/s12971-016-0107-6

Cokkinides, V. E., M. T. Halpern, et al. (2008). "Racial and Ethnic Disparities in Smoking-Cessation Interventions." American Journal of Preventive Medicine 34(5): 404-412.

Fiore MC, Jaén CR, Baker TB, et al. (2008). Treating Tobacco Use and Dependence: 2008 Update. Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services. Public Health Service. Retrieved from https://www.ahrq.gov/prevention/guidelines/tobacco/index.html.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Behavioral Health : Alcohol, Substance Use/Abuse

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

Primary Prevention, Screening

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

Children

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

https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measures/2019_Measure_402_MIPSCQM.pdf

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

This is not an eMeasure **Attachment:**

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

No data dictionary **Attachment:**

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

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

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

Not an instrument-based measure

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

No

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

N/A

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

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

Patients who were screened for tobacco use at least once within 18 months (during the measurement period or the six months prior to the measurement period) AND who received tobacco cessation counseling intervention if identified as a tobacco user.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

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

Documentation that the adolescent is currently not a tobacco user

OR

Documentation that the adolescent is a tobacco user AND any of the following:

-Advice given to quit smoking or tobacco use

-Counseling on the benefits of quitting smoking or tobacco use (e.g., "5-A" Framework)

-Assistance with or referral to external smoking or tobacco cessation support programs (e.g., telephone counseling 'quit line')

-Current enrollment in smoking or tobacco use cessation program

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

All patients aged 12-20 years with a visit during the measurement period

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

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

Patient Age is 12-20 Years on Date of Encounter

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

N/A

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

N/A

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

N/A

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

1. Start with Denominator

2. Check Patient Age:

a. If Patients Age is 12-20 Years on Date of Encounter equals No during the measurement period, do not include in Eligible Population. Stop Processing.

b. If Patients Age is 12-20 Years on Date of Encounter equals Yes during the measurement period, proceed to check Encounter Performed.

3. Check Encounter Performed:

a. If Encounter as Listed in the Denominator equals No, do not include in Eligible Population. Stop Processing.

b. If Encounter as Listed in the Denominator equals Yes, include in Eligible Population.

4. Denominator Population:

a. Denominator Population is all Eligible Patients in the Denominator.

5. Start Numerator

6. Check Patient Documented as Tobacco User AND Received Tobacco Cessation Intervention if Identified as a Tobacco User:

a. If Patient Documented as Tobacco User AND Received Tobacco Cessation Intervention if identified as a Tobacco User equals Yes, include in Data Completeness Met and Performance Met.

b. If Patient Documented as Tobacco User AND Received Tobacco Cessation Intervention if identified as a Tobacco User equals No, proceed to check Currently a Tobacco Non-User.

7. Check Currently a Tobacco Non-User:

a. If Currently a Tobacco Non-User equals Yes, include in Data Completeness Met and Performance Met.

b. If Currently a Tobacco Non-User equals No, proceed to check Tobacco Assessment or Tobacco Cessation Intervention Not Performed, Reason Not Given.

8. Check Tobacco Assessment or Tobacco Cessation Intervention Not Performed, Reason Not Given:

a. If Tobacco Assessment or Tobacco Cessation Intervention Not Performed, Reason Not Given equals Yes, include in the Data Completeness Met and Performance Not Met.

b. If Tobacco Assessment or Tobacco Cessation Intervention Not Performed, Reason Not Given equals No, proceed to check Data Completeness Not Met.

9. Check Data Completeness Not Met:

a. If Data Completeness Not Met, the Quality Data Code or equivalent was not submitted

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Claims, Electronic Health Records

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

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

This measure is currently in use as a MIPS Clinical Quality Measure under the Quality Payment Program. The source of the data are electronic health records. The data are collected using E.H.R data abstraction, and is reported by individual clinicians, groups, or third-party intermediaries on behalf of individual clinicians or groups.

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

No data collection instrument provided

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

Clinician : Group/Practice

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

Outpatient Services

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** NQF # 2803
Measure Title: Tobacco Use and Help with Quitting Among Adolescents
Date of Submission: 1/1/2020

Type of Measure:

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

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

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

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

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

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

2020 Submission

Testing was performed using data reported to the Centers for Medicare & Medicaid Services (CMS) Merit-Based Incentive (MIPS) Program

2015 Submission N/A

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

2020 Submission

Testing was performed using 2017 data.

2015 Submission

Testing of data element reliability and validity was performed using data from two study groups. In study group 1, data were obtained for care occurring from October 1, 2010 to December 31, 2011 (a 15-month observation period). For study group 2, data were obtained for care occurring between October 1, 2010 and March 2012 (an 18-month observation period).

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

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

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

2020 Submission

Group/Practice: 512 practices

2015 Submission

We field tested this measure in three pediatric centers. The participating sites included pediatric clinics affiliated with a children's hospital (this sample was selected from adolescents enrolled in Medicaid); a network of clinics serving homeless and vulnerable adolescents, and an adolescent medicine clinic affiliated with a children's hospital (which primarily provides behavioral health and gynecology care to young women). The participating sites were in different states and used different EHR vendors.

In addition, we tested for face validity using advisory panels, which included experts in measures development, adolescent medicine, and quality improvement (i.e. individuals well positioned to speak to a measure's face validity). See Submission form for the names and affiliations of panel members.

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

2020 Submission

Below are mean, median and ranges of denominator sizes (i.e., patients) included in the measure results.

Tobacco Use and Help with Quitting Among Adolescents: Denominator Size

Reporting level	N	Mean (%)	Min	25th	50th	75th	90th	Max	Interquartile Range
Group/practice	512	792	20	129	335	833	1822	22633	704

2015 Submission

Potentially eligible adolescents were 12 to 19 years old as of December 31, 2010 (thus adolescents in the study ranged from age 12 to age 20) and had at least one visit to the same primary care office or adolescent medicine clinic in both 2010 and 2011.

A total of 597 adolescents comprised the final study group. Site personnel assigned site-specific identification numbers to protect the confidentiality of the adolescents' records and maintained a crosswalk with the patient identifiers. The mean age of the sample was 15.5 years (range: 12 to 19 years). Slightly more than two-thirds of the sample was female (68.2%). African-American adolescents represented the largest proportion of the overall sample (44.4%) followed by non-Hispanic whites (30%). Approximately 93% of adolescents lived in households where English was the preferred language spoken at home.

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

2020 Submission: N/A

2015 Submission

Reliability Testing

The sample for reliability testing was a sub-sample of 75 adolescents from the initial sample of 597 (25 from each of the three sites).

Validity Testing

For validity testing, we compared performance against well-care visits. Thus, we used the sample from two of our sites; site 3 was excluded because it was an adolescent medicine clinic that served primarily female adolescents for behavioral health and gynecology care. The resulting sample was 400 adolescents from the initial sample.

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

2020 Submission

We did not assess rates by social risk factors. Social risk data are not available in this dataset.

2015 Submission

Adolescent's health insurance coverage (commercial, Medicaid, self-pay/other) was used as a proxy measure of family socioeconomic status.

2a2. RELIABILITY TESTING

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

2020 Submission

To assess reliability, we utilized the Beta-binomial model (Adams 2009) at the group/practice level 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 in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS measures. 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 accountable entities).

In addition to the point estimate of signal-to-noise reliability, NCQA will also provide the standard error and 95% confidence interval (95% CI) by June 2, 2020. NCQA will also include a summary of the methodology that was used to estimate the standard error and 95% CI.

Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

2015 Submission

Reliability was tested using two methods: inter-rater reliability and manual review/EHR extract comparison.

To assess inter-rater reliability, two reviewers independently collected data on 75 patients. Inter-rater reliability assessed the level of agreement for data elements related to tobacco cessation assistance between two independent abstractors reviewing the same data from the same data source. Agreement between abstractors was measured using the kappa statistic, which is a measure of agreement adjusted for agreement that could occur by chance. Kappa coefficients greater than 0.75 are indicative of excellent agreement.

In the manual review/EHR comparison, we assessed the agreement between rates calculated using manual EHR review compared to rates calculated automatically through an EHR extract. We report the kappa statistic here as well.

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

2020 Submission

Signal-to-noise reliability at the group/practice level is 0.996. NCQA will provide the standard error and 95% confidence interval by June 2, 2020.

2015 Submission

Inter-Rater Reliability

Table 1 presents the levels of agreement between the two manual reviewers together for the data elements of tobacco cessation assistance.

Table 1. Inter-rater Reliability of Manual Reviews for *Tobacco Use and Help With Quitting Among Adolescents* Data Elements[1]

	TOTAL	
Data elements	**Kappa Coefficient**	**95% Confidence Interval[2]**
Smoking Status as defined in CMS EHR Meaningful Use objectives	0.94	0.87, 1.00
Current tobacco use	n/a	n/a
Documentation of advice to quit smoking/using tobacco	n/a	n/a
Counseling on the benefits of quitting smoking/using tobacco	n/a	n/a
Referral to smoking/tobacco cessation support program	n/a	n/a
Enrolled in a smoking/tobacco use program	n/a	n/a

[1] Based on n=75 repeated ratings by two manual reviewers

[2] 95% confidence intervals listed as n/a are because neither rater could find any data available in these charts for those data elements, though in these cases percent agreement can be considered 100%

Comparison between manual review and automated EHR extract

Table 2 compares information on tobacco use documentation and help with quitting calculated from manual EHR review versus automated EHR data extracts for the same sample of adolescents.

Table 2. Agreement between Manual EHR Review and Automated EHR Extract: Information on Tobacco (N=597)

	Manual EHR Review	**Automated Data Extract**	**Kappa Coefficient**	**95% Confidence Interval**
Percentage of Adolescents with Tobacco Status Documented	70.9%	53.9%	0.52	0.45, 0.58
Percentage of Adolescents Who Are Current Tobacco Users	13.6%	7.7%	0.66	0.56, 0.76
Percentage of Adolescents Whose Tobacco Use Is Documented and Who Received Help With Quitting If They Are Users	61.6%	47.4%	0.52	0.45, 0.59

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

2020 Submission

The signal-to-noise reliability estimate is greater than 0.7, indicating the measure has very good reliability and provides confidence that one can distinguish the performance of one group/practice from another.

2015 Submission

Inter-rater Reliability

The agreement between the two reviewers was high for a large proportion of data elements (approximately 200). While we are reporting on our testing for the tobacco use measure, the variables we assessed in our

overall field test also included aspects of care related to demographics, sexual activity, chlamydia screening, depression screening, vaccinations, and other common well-care visit items.

There was high agreement for smoking status based on the *Meaningful Use* definition data element that was included in our measure (Kappa coefficient =0.94). The kappa coefficients for the remaining data elements could not be calculated because there was no variance in the ratings of either reviewer, primarily because the data elements were not documented.

Comparison between manual review and automated EHR extract

Overall, there was documentation of tobacco status for 70.9% of adolescents in the manual reviews compared to 53.9% in the automated extracts; the Kappa score (0.52) shows moderate agreement. In the manual reviews, 13.6% of adolescents were identified as smokers compared to 7.7% in the automated extract; the agreement is substantial (Kappa=0.66). For the proposed measure, the percentage of adolescents whose tobacco use is documented and who received help with quitting if they are users was 61.6% in the manual review versus 47.4% in the automated extract; the agreement was moderate (Kappa=0.52). There were substantial variations by site both in the results and the agreement between manual review and automated EHR extract.

2b1. VALIDITY TESTING

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

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

☐ **Performance measure score**

☒ **Empirical validity testing**

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

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

2020 Submission

Empiric Validity Testing of Performance Measure Score

We tested for construct validity by exploring whether the *Tobacco Use and Help with Quitting Among Adolescents* measure was correlated to the *Unhealthy Alcohol Use: Screening and Brief Counseling* measure. Both measures assess screening and follow-up of substance use. We hypothesized that reporting entities that perform well on the *Tobacco Use* measure should perform well on the *Alcohol Screening* measure. While there were three other preventive care and screening measures that may have been relevant for construct validity, unfortunately, these measures did not have sufficient sample size to conduct testing (n < 5 reporting entities).

To test these correlations, we used a Pearson correlation test. This test estimates the strength of the linear association between two variables. The magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

Systematic Assessment of Face Validity of Performance Measure Score

Validity refers to whether the measure represents the concept being evaluated. During development, our team reviewed the specifications and field test results with our advisory panels, which included individuals well positioned to speak to a measure's face validity. We convened a multi-stakeholder advisory panel with representation from a wide range of stakeholders, including consumers, pediatricians, family physicians, adolescent medicine physicians, health plans, state Medicaid agencies and researchers. In addition, we convened three targeted panels of stakeholders with particular relevance to the measures: we partnered with the National Partnership for Women and Families to convene a panel of consumer and family advocates; we partnered with the American Academy of Pediatrics to convene a panel of pediatricians, including adolescent medicine physicians; and we convened a panel of state Medicaid and CHIP representatives.

2015 Submission

Method of assessing face validity

Validity refers to whether the measure represents the concept being evaluated. To assess different perspectives on the measure's face validity, NCINQ reviewed the specifications and field test results with our advisory panels, which included individuals well positioned to speak to a measure's face validity. We convened a multi-stakeholder advisory panel with representation from a wide range of stakeholders, including consumers, pediatricians, family physicians, adolescent medicine physicians, health plans, state Medicaid agencies and researchers. In addition, we convened three targeted panels of stakeholders with particular relevance to the measures: we partnered with the National Partnership for Women and Families to convene a panel of consumer and family advocates; we partnered with the American Academy of Pediatrics to convene a panel of pediatricians, including adolescent medicine physicians; and we convened a panel of state Medicaid and CHIP representatives.

Method of assessing known groups validity

While any clinical encounter with adolescents, including sports physicals or acute care visits, represents an opportunity to discuss risky behaviors, designated well-care visits provide an important opportunity for these conversations. For this reason, NCINQ chose to evaluate the known-groups validity, defined as the ability of the measure to meaningfully differentiate distinct groups, by comparing the performance rates of adolescents who did not have any well-care visits in the measurement period to those who had one or more well-care visits. The manual reviewers abstracted the total number of well-care visits that were completed from October 1, 2010 to December 31, 2011. We defined well-care visits based on diagnosis or procedures codes or a visit that included documentation of health and developmental history, a physical exam, and health education/anticipatory guidance. The total number of well-care visits was transformed into a dichotomous variable to indicate whether the adolescent had any well-care visits (yes/no). NCINQ excluded Site 2 from the known groups validity analysis; this site is an adolescent medicine clinic that served primarily female adolescents for behavioral health and gynecology care.

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

2020 Submission

Empirical Validity Testing of Performance Measure Score
The Pearson correlation coefficient for the *Tobacco Use and Help with Quitting Among Adolescents* and *Unhealthy Alcohol Use: Screening and Brief Counseling* measures was as follows:

Group/Practice level: 0.37 ($p < 0.0001$)

For face validity results, see 2015 submission information below.

2015 Submission

Face Validity

Our advisory panels concluded the measure is a valid way to assess tobacco status, use and follow-up among adolescents.

Known Groups Validity

As shown in Table 3, documentation of tobacco use and help with quitting was significantly higher among adolescents who had at least one well-care visit in the measurement period compared to adolescents without designated well care visits. The results were significant (p-value <0.0001).

Table 3. Known Groups Validation: Tobacco Use and Help with Quitting Among Adolescents with and Without Designated Well Care Visits[1]

	Had 1 or More Well-Care Visits in Measurement Period		
Percentage of adolescents whose tobacco use is documented and who received help with quitting if they are users	Yes	No	*p*-value
Sites 1 and 3 (combined)	58.9%	39.2%	<0.0001

[1] Data from EHR manual review (N=400)

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

2020 Submission

For the purposes of this analysis and the intended use of this measure to evaluate the quality of care for adolescents across practices, correlation was considered high (strong) if the correlation coefficient is 0.75 to 1, moderate if 0.25 to 0.75, and low (weak) if 0 to 0.25.

The correlation value of 0.37 is moderate, suggesting that reporting entities that performed well on the *Tobacco Use* measure are moderately likely to perform well on the *Unhealthy Alcohol Use* measure.

2015 Submission

Our advisory panels concluded the measure as specified is a valid way to assess tobacco status, use and follow-up in adolescents. Our interpretation of these results is that the measure has sufficient face validity.

Known Groups Validity

Documentation of tobacco use and help with quitting was significantly higher among adolescents who had at least one well-care visit in the measurement period compared to adolescents without designated well care visits. Based on these results, we conclude the measure shows what we would expect from measure performance among these two groups. We expect patients with more well care visits to be compliant for the tobacco-use measure, and our results aligned with this expectation.

2b2. EXCLUSIONS ANALYSIS

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

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

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

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

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

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

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

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

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

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

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

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

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

Provide the statistical results from testing the approach to controlling for differences in patient characteristics

(case mix) below.
If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)
2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

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

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

2020 Submission

To demonstrate meaningful differences in performance, we calculated 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, we calculated an independent sample t-test of the performance difference between two randomly selected group practices at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each practice. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than 0.05, then the two practices' performance is significantly different from each other.

2015 Submission

Our sample of three sites did not provide sufficient data to conduct statistical tests. We provide information on the mean rates across the three sites below.

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

2020 Submission

Tobacco Use and Help with Quitting Among Adolescents

Reporting level	N	Mean eligible population	Mean (%)	Min	25th	50th	75th	90th	Max	IQR	p-value
Group/ practice	512	792	89.1	4	87	96	99	100	100	12	p < 0.0001

IQR: Interquartile Range

p-value: p-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

2015 Submission

Performance rates for the Tobacco Use and Help with Quitting Among Adolescents measure based on manual EHR review are presented by site and total sample in the table below. The overall rate was 61.6%. Rates vary from 44.5 percent documentation to 85.3 percent.

Performance Rates for Tobacco Use and Help with Quitting among Adolescents in Manual EHR Review, Overall and by Site

	Overall	Site 1	Site 2	Site 3
Percentage of Adolescents whose tobacco use is documented and who received help with quitting if they are users	61.6%	55.5%	85.3%	44.5%

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

2020 Submission

For groups/practices, the IQR was 12%. This gap represents an average of 95 adolescents in high-performing versus low-performing entities.

2015 Submission

Rates varied from a low of 44.5 percent documentation to a high of 85.3 percent documentation. While site-to-site variation can be explained, in part, by differences in the availability of data elements, content of free-text notes, and site characteristics, we believe that variations in these results would imply variations among providers.

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

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

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

2020 Submission

We did not conduct a missing data analysis.

2015 Submission

This measure is collected with a complete sample, there is no missing data on the overall measure.

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

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

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

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

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

Some data elements are in defined fields in electronic sources

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

This measure uses the Clinical Quality Measure reporting method. Some components of this measure are available in structured fields, while others are available in narrative notes or other non-structured fields.

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

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

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

Our testing revealed that this measure was feasible for clinicians to report. No known difficulties were identified with this measure's operational use as a PQRS and subsequently MIPS measure.

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

Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting CMS Merit-Based Incentive Payment System (data reported publicly via Physician Compare) Program website: https://qpp.cms.gov/mips/overview Publicly reported data: https://data.medicare.gov/Physician-Compare/Physician-Compare-2017-Individual-EC-Public-Report/93x5-v9gz CMS Merit-Based Incentive Payment System (data reported publicly via Physician Compare) Program website: https://qpp.cms.gov/mips/overview Publicly reported data: https://data.medicare.gov/Physician-Compare/Physician-Compare-2017-Individual-EC-Public-Report/93x5-v9gz Payment Program CMS Merit-Based Incentive Payment System https://qpp.cms.gov/mips/overview CMS Merit-Based Incentive Payment System https://qpp.cms.gov/mips/overview

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

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

CMS QUALITY PAYMENT PROGRAM: This measure is used in the Quality Payment Program (QPP) which is a quality and cost incentive program that uses payment adjustments to promote high quality and high value care delivery by eligible clinicians (EC). QPP provides performance-based payment adjustments to ECs, both negative and positive, for services furnished to Medicare Part B beneficiaries. EC performance is graded on quality measure performance, cost of care, engagement in clinical practice improvement activities, and use of Certified EHR Technology (CEHRT). Performance can be reported at the individual (clinician) or group (practice) level. In 2017, 1,006,319 ECs participated in MIPS, representing 95% of all eligible clinicians across the 50 states. 54% participated as a part of a group, 12% as individual clinicians, and 34% as a part of an Advanced Payment Model.

References:

Centers for Medicare and Medicaid Services (CMS). (2018). 2017 Quality Payment Program Reporting Experience. Baltimore, MD: Centers for Medicare and Medicaid Services. Retrieved from: https://www.pcpcc.org/resource/2017-quality-payment-program-reporting-experience

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

N/A - this measure is publicly reported through the Merit-Based Incentive Payment System (MIPS).

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

N/A – this measure is publicly reported through the Merit-Based Incentive Payment System (MIPS).

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

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

When clinicians and groups report their performance on this measure for MIPS, CMS provides them with feedback reports to inform performance improvement efforts. All individual and group performance data is reported publicly through Physician Compare, and annual benchmarks are publicly available to enable clinicians to understand how their performance compares to national benchmarks.

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

Feedback reports are made available to clinicians in July following the measurement year (i.e. feedback reports for 2017 were available in July of 2018). Reports include performance rates as well as the associated payment adjustment to Medicare Part B payments. The full performance data set became available on Physician Compare in 2019. The Physician Compare data set includes measure performance scores for all individual clinicians and groups that reported measures to MIPS.

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

Describe how feedback was obtained.

NCQA-developed measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels and posting for public comment.

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

This measure was posted for public comment in 2015. Overall, the comments were generally appreciative of this measure's intent to address a performance gap in an area of importance to adolescent health. Commenters highlighted the importance of including electronic cigarettes in the measure. We received further comments suggesting that this measure is duplicative of other existing measures which look at tobacco use in adults.

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

We have not received any additional feedback on this measure.

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

No feedback has been received that indicated need for modification.

Improvement

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

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

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

2017 is the first and only year that publicly reported MIPS data is available on Physician Compare for this measure. Thus, we are unable to describe demonstrated performance improvement year over year.

Clinicians and groups are able to use performance results to inform quality improvement efforts related to screening and counseling adolescents for tobacco use. MIPS performance results impact physician payment and are publicly reported and accessible by patients and families. Therefore, clinicians are incentivized both financially and reputationally to perform well on this measure. When results indicate low rates of screening and counseling, clinicians can implement processes to improve screening practices and documentation of tobacco use status.

4b2. Unintended Consequences

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

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

No unintended consequences were identified for this measure.

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

Unexpected benefits could include better, more transparent, provider-patient communication about risky behaviors beyond tobacco use. Providers consistently engaging in conversations with adolescents about tobacco use might naturally open the door to more honest discussions of risky youth behaviors, such as substance use or risky sexual behavior, and opportunities to counsel on risk avoidance/mitigation.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0028 : Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

No

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

While both measures look at tobacco use and cessation counseling, they assess different target populations. NQF 0028 measures tobacco use in adults aged 18 and older. This measure assesses tobacco use in adolescents who are between the ages of 12 and 20. The expected impact on data collection burden due to the overlapping age range of 18-20 is minimal.

5b. Competing Measures

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

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

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

N/A – the measures assess different target populations.

Appendix

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

No appendix Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance

Co.2 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-3500-

Co.3 Measure Developer if different from Measure Steward: National Committee for Quality Assurance

Co.4 Point of Contact: Brittany, Wade, wade@ncqa.org, 202-530-0463-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

National Collaborative for Innovation in Quality Measurement (NCINQ) Consumer Panel

Joan Alker, MPhil, Georgetown Center for Children and Families

Roni Christopher, MEd, OTR/L, PCMH-CCE, The Greater Cincinnati Health Collaborative

Daniel Coury, MD, Nationwide Children's Hospital

Eileen Forlenza, Colorado Medical Home Initiative, Children and Youth with Special Health Care Needs Unit

Michaelle Gady, JD, Families USA
Janis Guerney, JD, Family Voices
Jocelyn Guyer, MPA, Georgetown Center for Children and Families
Catherine Hess, MSW, National Academy for State Health Policy
Carolyn Muller, RN, Montgomery County Health Department
Cindy Pellegrini, March of Dimes
Judith Shaw, EdD, MPH, RN, VCHIP
Stuart Spielman, JD, LLM, Autism Speaks
Michelle Sternthal, PhD, March of Dimes
NCINQ Measurement Advisory Panel
Mary Applegate, MD, Ohio Department of Job and Family Services
Katie Brookler, Colorado Department of Health Care Policy and Financing
Cathy Caldwell, MPH, Alabama Department of Public Health
Ted Ganiats, MD, University of California, San Diego
Darcy Gruttadaro, JD, National Allegiance on Mental Illness
Jennifer Havens, MD, NYU School of Medicine
Virginia Moyer, MD, MPH, FAAP, Baylor College of Medicine, USPSTF
Edward Schor, MD, Lucile Packard Foundation for Children's Health
Xavier Sevilla, MD, FAAP, Whole Child Pediatrics
Gwen Smith, Illinois Department of Healthcare and Family Services/Health Management Associates
Janet (Jessie) Sullivan, MD, Hudson Health Plan
Kalahn Taylor-Clark, PhD, MPH, George Mason University
Craig Thiele, MD, CareSource
Jeb Weisman, PhD, Children's Health Fund
Charles Wibbelsman, MD, Kaiser Permanente Medical Group, Inc.
NCINQ Clinician Advisory Panel
Elizabeth Alderman, MD, FAAP, Albert Einstein College of Medicine
Sarah Brewington, MD, Sandhills Pediatrics Inc
Gale Burstein, MD, MPH, FAAP, FSAHM, Women and Children's Hospital of Buffalo, NY
Barry Bzostek, MD, FAAP, Women and Children's Hospital of Buffalo, NY
Danielle Casher, MD, FAAP, St. Christoper's Hospital for Children
Edward Curry, MD, FAAP, Emergency Department, St. Christopher's Hospital for Children, PA
Eve Kimball, MD, FAAP, Southern California Permanente Medical Group
Paul Melinkovich, MD, FAAP, Kaiser Permanente
Jackie Nelson, MD, FAAP, Lander Regional H
Ellen Squire, MD, FAAP, HaysMed Pediatric Center
NCINQ State Panel
Mary Applegate, MD, Ohio Department of Job and Family Services
Sharon Carte, MHS, State of West Virginia Children's Health Insurance Program

Susan Castellano, Minnesota Department of Human Services

Catherine Hess, MSW, National Academy for State Health Policy

Michael Hogan, PhD, New York State office of Mental Health

Barbara Lantz, MN, RN, State of Washington Department of Social and Health Services, Medicaid Purchasing Administration

Judy Mohr Peterson, PhD, Oregon Health Authority

Tracy Plouck, MPA, Ohio Department of Mental Health

Gina Robinson, Colorado Department of Health Care Policy and Financing

Janet Stover, Illinois Association of Rehabilitation Facilities

Eric Trupin, PhD, University of Washington

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2015

Ad.3 Month and Year of most recent revision: 04, 2015

Ad.4 What is your frequency for review/update of this measure? Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

Ad.5 When is the next scheduled review/update for this measure? 12, 2021

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. THE MEASURSE AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.