NATIONAL QUALITY FORUM
Driving measurable health improvements together

Prevention and Population Health Fall 2022 Submissions Cycle: Pre-evaluation Comments

Contents

Funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001

NQF #0028 and Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention (National Committee for Quality Assurance)

Pre-evaluation Standing Committee Comments

1a. Evidence

- No comments.
- Established with precedent.
- I agree that there is no need for repeated discussion and a vote on evidence. Solid evidence of relationship between measure and smoking cessation intermediate outcome. Evidence is directly applicable to the process of care being measured.
- This is a maintenance measure with a logic model indicating that screening and the offering of cessation does result in smoking cessation. New evidence is not offered.
- No new studies since most recent USPTF systematic review and guidance. Process relates to desired outcomes (reduce heart & lung disease, stroke).
- There is strong evidence that tobacco screening and cessation counseling and pharmacotherapy can have positive impacts on patients. The measurement period is not clearly articulated.
- Based on USPSTF systematic review and recommendation.
- The evidence relates to the specific process being measured by showing the link between receiving a cessation intervention and quitting among patients who report tobacco use. The systematic reviews and trials provide evidence of the effectiveness of cessation interventions in helping patients quit smoking, while the 2021 Final Recommendation Statement from the USPSTF supports the use of specific interventions (behavioral and FDA-approved pharmacotherapy) for tobacco smoking cessation in nonpregnant and pregnant adults. The evidence is directly related to the process being measured. It supports the use of cessation interventions in helping patients quit smoking and highlights the specific interventions that can be used. The direct link between the evidence and the measure shows its relevance and importance. The process being measured are directly related to the desired outcome. The structure of having cessation interventions as part of the healthcare system, the process of delivering these interventions to patients who report tobacco use, and the outcome of increased quitting rates among patients who receive these interventions, are all steps towards the desired outcome of reducing tobacco use. Based on my knowledge, there is no available study or information that demonstrates any alteration in the evidence presented by the developer.
- This is a process measure. The evidence is rated high. The developer provided new evidence from systematic reviews completed between 2013 and 2022 as well as randomized and non-randomized trials show a link between receiving a cessation intervention and quitting among patients who report tobacco use. The developer used the 2021 Final Recommendation Statement from the USPSTF to gather the following evidence: Clinicians should ask all adults, regardless of pregnancy status, about tobacco use. The net benefit of behavioral and FDA-approved pharmacotherapy interventions, either alone or combined, for tobacco smoking cessation in nonpregnant adults is substantial. The net benefit of behavioral interventions for

tobacco smoking cessation in pregnant adults is substantial. Evidence regarding the benefits and harms of pharmacotherapy for tobacco cessation among pregnant adults is insufficient. Evidence regarding the benefits and harms of using e-cigarettes for tobacco cessation among pregnant and nonpregnant adults is insufficient.

- Evidence for this measure remains strong and unchanged since the last review and should be retained.

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

- Yes, gaps in care in smoking cessation counseling and in smoking-related morbidity and mortality. Studies have consistently shown that members of some racial or ethnic groups receive less counseling from their doctors and have poorer outcomes than White patients.
- Evidence supports the gap and disparities.
- Developer provided evidence of ongoing gap in care, and evidence of various sociodemographic disparities.
- Disparity data is not available on this measure. Given higher rates of morbidity and mortality amongst BIPOC communities, disparity data would be helpful.
- Performance gap data were provided and indicate a gap in care. Disparities data was provided and indicates disparities in care.
- Data are not as current as desirable, but that is not the fault of the measure developers. The most recent data demonstrate performance gaps.
- The developer reports a mean performance of 71% with a standard deviation of 30%, presumably based on the proposed measure, although this is not stated explicitly. The developer also cites a variety of other measures from the National Health Interview Survey to demonstrate gaps and disparities, but these statistics are not the same as the proposed measure, and the data are from 2015.
- Yes, current performance data on the measure is provided. The developer provided data from the Merit-based Incentive Payment System for 2020 which showed a mean performance of 70.71%. The developer also looked at data from the 2015 National Health Interview Survey which showed that 68% of adults wanted to stop smoking and 55.4% made a quit attempt in the past year. Disparities data is not yet available for analysis and reporting. The provided data shows a mean performance of 70.71 percent with a standard deviation of 29.93, which demonstrates variability in performance. This suggests that there is room for improvement and a gap in care. Data on disparities for the measure is provided from the 2015 National Health Interview Survey. The data shows differences in the percentage of individuals advised to quit smoking by age, race, insurance status, disability, and serious psychological distress. The data from the 2015 National Health Interview Survey demonstrates disparities in care by showing differences in the proportion of different demographic groups who were advised to quit smoking by a health professional. It showed that individuals 65 years and older, those between 45-64 years old, those with disabilities, and those with serious psychological distress were more likely to be advised to quit smoking compared to younger individuals, those without disabilities, and those without serious psychological distress. The data also showed that white individuals were more likely to be advised to quit compared to Hispanic, Indian/Alaska Native, and Asian individuals. Smokers with insurance were more likely to receive advice to quit compared to uninsured smokers. These disparities suggest that there may be a gap in care for certain

demographic groups.

- The developer provided 2020 (January 1-December 31, 2020) opt-in data from Merit-based Incentive Payment System which included 19,427 physicians and other clinicians (e.g. nurse practitioners, Physician Assistants). The developer found a mean performance of 70.71 percent with a standard deviation of 29.93. The developer also looked at data from the 2015 National Health Interview Survey which analyzed the percentage of adults who attempted to quit and who received a cessation intervention. The developer stated although this measure is included in federal reporting programs, disparities data has not yet been made available for analysis and reporting.
- There is moderate opportunity for improvement. Gaps were noted for age, ethnicity, race, insurance status, disabilities, and psychological distress.

2a1. Reliability - Specifications

- No concerns.
- No concerns.
- Measure specifications are clear and precise.
- No concerns. There is new reliability testing.
- No concerns.
- To measure counseling either/or pharmacotherapy is reliably captured. However, what is not reliably captured is if there are more counseling services, pharmacotherapy, or combined. A "yes/no" about counseling and/or pharmacotherapy limits the ability to determine the impact of different approaches with patients.
- Calculating the measure requires ascertaining whether (1) patients are screened for tobacco use at least once during the measurement period and (2) those found to be using tobacco receive an appropriate tobacco cessation intervention. The application, however, does not mention how these events are ascertained, and this important dimension of reliability is not addressed in the application, and the reliability testing does not get at this. The application also does not address the appropriateness of the screening or the interventions, or how they are ascertained from the available records. The fundamental question is whether the calculated rates truly represent whether patients have received the screening and interventions indicated by the USPSTF recommendation. Presumably this was assessed and found to be ok in an earlier application.
- The measure specifications have remained unchanged since the last review.
- The developer conducted new reliability testing. The developer performed a signal-to-noise reliability test for each population (1: those screened for tobacco use, 2: those who received an intervention, and 3: those screened and who received an intervention). The data is from 19,427 individual physicians and other clinicians who opted-in to the Merit-based Incentive Payment System (MIPS). For population 1, scores ranged from 0.888 to 1 with a mean reliability of 0.994. The 25th and 75th percentiles were 0.995 and 0.999 respectively, with a standard deviation of 0.012. For population 2, scores ranged from 0.887 to 1 with a mean reliability of 0.992. The 25th and 75th percentiles were 0.993 and 0.998 respectively, with a standard deviation of 0.014. For population 3, scores ranged from 0.896 to 1 with a mean reliability of 0.994. The 25th and 75th percentiles were 0.963 and 0.994 respectively, with a standard deviation of 0.023. The

developer attests the results indicate very good reliability, and that the variation is caused by real differences in performance across reporting entities.

- Measure specifications have not changed since the last review. The developer did conduct new reliability testing. Reliability appears to be very high for each of the three populations. No concerns identified.

2a2. Reliability - Testing

- No concerns.
- No concerns.
- Developer conducted new reliability testing, and attests the results indicate very good reliability, and that the variation is caused by real differences in performance across reporting entities.
- No concerns.
- No.
- No.
- The proposed measure is a kind of composite that addresses two concepts: (1) patients are screened for tobacco use and (2) those found to be using tobacco receive a tobacco cessation intervention. The application specifies the numerator and denominator for three "populations," but only the third is the same as the proposed measure, i.e. patients are screened and those found to be tobacco users receive an intervention. The first "population" addresses the first component of the measure, i.e., patients are screened. The second "population" is essentially the proportion of those who screen positive who receive an intervention. Thus, the tests of the first and second "population" are not relevant to the assessment of the reliability of the proposed measure, and my rating is based only on the test of the third "population." The proposed measure is intended to be used on the individual provider level, and reliability testing is appropriately conducted using the beta-binomial method based on a sample of over 19,000 individual clinicians in the Merit-based Incentive Payment System (MIPS). The results seem suspiciously good, with a mean reliability of 0.994, a standard deviation of 0.023, and a low score of 0.963. It is difficult to believe that none of the more than 19,000 clinicians were less reliable than 0.963. Furthermore, the results for the three populations are suspiciously close to one another, even though the concepts are so different. Consequently, I wonder whether the calculations were done properly. Because of my concerns about the lack of information on specifications, and about the testing calculations, I would say that the data on reliability are Insufficient.
- No concerns about reliability of the measure. The test results indicate high reliability and variation is likely due to real differences in performance.
- No concerns. Specifications are precise, unambiguous, and complete.
- None noted.

2b1. Validity Testing

- No concerns.

- No concerns.
- No. Moderate correlations with alcohol screening results.
- No concerns. There is new validity testing.
- No.
- Similar to concerns raised above about reliability.
- No concerns.
- No concerns found in the testing results.
- No concerns. Moderate level. The developer performed a Pearson correlation test for construct validity to determine if the tobacco measure results correlated with another behavioral health screening measure, specifically the Preventive Care and Screening: Unhealthy Alcohol Use: Screening and Brief Counseling. The developer noted a difference between the number of rates for the alcohol and tobacco measures. The alcohol measure has one rate assessing whether patients who were screened and identified as an unhealthy alcohol user received brief counseling, and the tobacco measure has three rates (1: those screened for tobacco use, 2: those who received an intervention, 3: those screened and who received an intervention). Each tobacco rate was assessed separately against the alcohol measure rate. For population 1, the developer reports the rate is positively and moderately associated with the alcohol measure rate. The correlation coefficient was 0.461, and a p value <0.001. For population 2, the rate is positively and moderately associated with the alcohol measure rate. The correlation coefficient was 0.371, with a p value <0.001. For population 3, the rate is positively and moderately associated with the alcohol measure rate. The correlation coefficient was 0.434, with a p value <0.001. The developer attests the tobacco measure performance is moderately associated with the alcohol screening measure. Therefore, the developer suggests that clinicians who perform well on one of these preventive behavioral health measures will likely perform well on the other.
- None noted.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)

- No comment.
- Evidence for consistent exclusions & adequate risk adjustment
- No exclusions or risk adjustment.
- The data is not risk adjusted or stratified.
- No exclusions. Exceptions are appropriate.
- Exclusions were not well-articulated. "Documentation of medical reasons…." leaves this a bit open-ended.
- NA
- The measure under review does not use any exclusions and is not risk-adjusted or stratified.
- The developer states they do not have information on the extent and distribution of missing data because this information is not collected by CMS.

- No risk adjustment using social risk factors was reported.

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)

- No.
- Differences in quality are demonstrated, results are comparable, no significant threats.
- The developer states they do not have information on the extent and distribution of missing data because this information is not collected by CMS.
- No information is available on missing data. No concerns.
- No concerns.
- To measure counseling either/or pharmacotherapy is a valid measure. However, what is not captured is if there are more counseling services, pharmacotherapy, or combined. A "yes/no" about counseling and/or pharmacotherapy limits the ability to determine the impact of different approaches with patients.
- One would imagine that clinical records sometimes do not indicate whether screening was conducted, the results of that screening, or whether an appropriate intervention was delivered. The developer's statement that "they do not have information on the extent and distribution of missing data because this information is not collected by CMS" does nothing to reassure me on this score. Consequently, I have serious concerns about this aspect of validity.
- The analyses indicate that the measure identifies meaningful differences about quality by calculating an IQR for each rate and then conducting an independent sample t-test to determine if the differences were statistically significant. The results showed a statistically significant difference in performance for all three populations, with p values <0.001. The information provided does not mention any specific comparison of results. The developer states they do not have information on the extent and distribution of missing data, so it is not possible to determine if missing data constitutes a threat to the validity of the measure.
- The developer states they do not have information on the extent and distribution of missing data because this information is not collected by CMS. The measure only uses one set of specifications for this measure.
- A Pearson correlation test was conducted with unhealthy alcohol use and was found to be moderately associated for the three populations and does identify differences in quality performance. No information offered for missing data.

3. Feasibility

- Required data elements available in EHRs.
- Feasibility quite evident.
- No concerns.
- No concerns about feasibility. This data has been collected and is operationalized.
- No concerns regarding feasibility.
- No concerns.

- No concerns.
- The data elements required for the measure are routinely generated and used during care delivery and are available in electronic form. There are no concerns regarding the data collection strategy, as it is carried out by healthcare personnel and claims data is captured electronically.
- Data elements are collected and used by healthcare personnel. Claims data is captured electronically with encounter codes for the denominator and CPT II codes for the numerator. However, the developer states registry implementation may vary. There is an eCQM version of this measure (CMS138), but it has been withdrawn from endorsement review because NCQA is not in a position to retest. The developer has not identified feasibility issues related to data collection, availability of data, missing data, sampling, or patient confidentiality.
- No feasibility issues were reported.

4a. Use

- Lacking feedback of performance results.
- Accountability & transparency aligned with feedback.
- No concerns.
- The data is publicly reported and used in an accountability program.
- Public reporting available through care compare. Feedback to those measured is available.
- No concerns.
- No concerns.
- The measure is publicly reported through the Quality Payment Program Merit-based Incentive Payment System, and also through the Health Resources and Services Administration Uniform Data System. The measure performance results are published annually on the CMS Physician Compare website, making the results available outside of the organizations or practices whose performance is measured. The measure under review is provided with performance results or data annually through CMS' Physician Compare website. However, the developer notes that CMS does not provide information regarding feedback from measured entities on measure performance and implementation. It is unclear whether those being measured have been given assistance with interpreting the measure results and data, or if they have been given an opportunity to provide feedback on the measure's performance or implementation. Whether the feedback has been considered when changes are incorporated into the measure is also unknown.
- The measure is currently in the Quality Payment Program Merit-based Incentive Payment System, which is a payment program and publicly shares submitted data and the Health Resources and Services Administration Uniform Data System, which provide a standardized reporting system for core set of information. The measure is also part of the Million Hearts Clinical Quality Measures, which are a focused set of high impact clinical quality measures for the ABCS (Aspirin when appropriate, Blood pressure control, Cholesterol management, and Smoking cessation) that are aligned across public and private national programs. CMS publishes measure performance results annually on its Physician Compare website. However, the developer notes that CMS does not provide information regarding feedback from measured

entities on measure performance and implementation.

- Performance on measure is reported back to those being assessed and have had an opportunity to provide feedback.

4a. Usability

- None.
- Use of administrative data facilitates the needed scale.
- What does the following comment in the staff review mean? "The developer did not report any improvement results due to limited availability of QPP data."
- I do not see any unintended consequences.
- Currently used in PI/VBP. The only unintended consequence I can foresee is relates to adult tobacco users who are reluctant to see medical providers who may avoid further encounters in response to tobacco use questions or counseling. But this risk exists for any population that is care-avoidant or preventive-care avoidant, so benefits outweigh risks.
- Related to 4b1, screening and counseling during the initial visit seems feasible and desirable.
- No concerns.
- The developer did not report any results regarding improvement or potential harm.
- No potential harms reported.
- No improvements, unexpected findings, potential harms were reported.

5: Related and Competing Measures

- No additional steps needed.
- Potential competing HEDIS measure?
- There are no known competing measures.
- No.
- None.
- No concerns.
- There is a related competing measure NQF #0027. However, the focus of the measures is slightly different.
- The developer states there are no endorsed related or competing measures but noted a related HEDIS measure titled Medical Assistance with Smoking and Tobacco Use Cessation [MSC]. The measure was endorsed by NQF under NQF #0027; however, endorsement was removed in 2020. Harmonization - This measure focuses on routine tobacco screening and tobacco cessation interventions with the intent of assessing clinician performance. NQF #0027 only assesses individuals who report smoking by capturing what percentage receive advice from a provider on quitting.
- No competing measures were reported.

Pre-evaluation Public and Member Comments

No comments were received.

NQF #0038 Childhood Immunization Status (National Committee for Quality Assurance)

Pre-evaluation Standing Committee Comments

1a. Evidence

- Measure aligned with guidelines.
- Evidence supports.
- I agree that there is no need for repeated discussion and a vote on evidence.
- This is a maintenance measure with evidence to support the measure focus.
- I found it difficult to follow the evidence-related responses to work backward to the evidence base. I understand the measure is based on ACIP recommendations and that ACIP uses an Evidence to Recommendations framework, but this seems opaque, as indicated by the submitted responses which cut and paste information from the ACIP website. When I attempted to locate a EtR for any vaccine included in the measure, I was unable to do so. Given the growing skepticism around vaccines, I find it concerning that the evidence supporting this measure is not widely accessible nor does it meet the usual requirements for grading of evidence.
- ACIP recommendations support the measures focus. With ACIP recommending the COVID vaccine for this age group, should that be added to this measure?
- The evidence for the measure is based on a systematic review and is related directly to the specific process being measured. It supports the intended relationship between the process and desired outcomes. No new studies or information that changes the evidence base for this measure have been cited in the submission.
- Based on up-to-date ACIP evidence review & recommendation.
- Evidence for this measure remains strong and unchanged since the last review and should be retained.
- The developer stated that they updated the citation of the clinical practice guideline to reflect the most recent recommended immunization schedule – provided a systematic literature review, though the evidence is not graded at the individual vaccine level. They also noted that there were no changes in clinical recommendations, therefore the measure continues to be aligned with the guideline. The developer stated that ACIP's recommendations for each vaccine are developed after reviews of vaccine-related data and that Evidence to Recommendation (EtR) framework is then used to summarize the key factors in the data. The developer noted that they were unable to find an overall rating of the evidence at the individual vaccine level. They further stated that ACIP linked to Evidence to Recommendation findings that detail out the rating of evidence for each question asked in the framework. The developers further stated that they did not find a summary of the net benefit and consistency at the vaccine level, but that the ACIP workgroup members assess the evidence and make recommendations based on findings of

immunogenicity and safety, expert opinion, and stakeholder input.

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

- Well-documented disparities in immunization rates for certain vaccines more than others in Medicaid populations. Developer did not provide disparities data, but CA Medicaid program demonstrated disparity in Black population for related CIS-3 measure.
- Gap & disparities evidence supports.
- In general, there is a very high level of performance on most of the measures, at the 80-90+% level. Worth discussion about level at which national performance measures are still required. Developer asserts but does not document disparities. Also worth discussion.
- There is a performance gap and disparity data can be visible at the health plan level.
- Performance data continue to indicate a gap in care. Difference in Medicaid plans vs. commercial plans implies disparities. No specific data were provided on disparities by subgroup.
- Performance gap data were provided. There continue to be performance gaps.
- Yes, current performance data on the measure was provided. The provided data demonstrates a gap in care for both commercial and Medicaid plans for different immunizations. The mean performance rate for the commercial plans is higher compared to the Medicaid plans for all the years, indicating an opportunity for improvement for Medicaid plans. The variability in the performance rate, as shown by the IQR and standard deviation, also highlights the gap in care. The developer has stated that the measure of disparity can be stratified by demographic variables such as race/ethnicity or socioeconomic status, but data to support this claim has not been provided.
- Medicaid vs. commercial could be regarded as evidence of disparities by socioeconomic position (SEP), but the developer does not explicitly make this case. Furthermore, in addition to the data provided, the developer should have provided summary data from the National Immunization Survey, which surely would have documented important disparities by race, ethnicity, SEP, etc.
- There is moderate opportunity for improvement. Measure can be stratified using demographic data elements. Gaps were noted for ethnicity, race, %FPL and insurance status; however, data to support this was not provided.
- The developer provided performance measure rates for each of the ten immunizations for the years 2019 – 2021 for both commercial and Medicaid plans using HEDIS data. For each vaccine type, there was one commercial and one Medicaid plan per year, resulting in three years of data for each plan type. A performance gap does exist for all vaccines for each year. Disparities: The developer stated that the measure can be stratified by demographic variables such as race/ethnicity or socioeconomic status to assess the presence of healthcare disparities if the data are available to a plan. The developer cited evidence from the National Immunization Survey which they stated showed that national coverage with most routine childhood vaccines remained stable, however, disparities in immunization coverage have been seen in uninsured, Black, and Hispanic patients, and patients living below the federal poverty line compared to individuals who were privately insured, White, or living at or above the federal poverty line. The developer did not provide data to support this claim.

2a1. Reliability – Specifications

- No concerns.
- Reliability adequate for population studies
- Specifications clear.
- No concerns. There is new reliability testing.
- No concerns. Can be consistently implemented.
- No concerns.
- No concerns about the likelihood that the measure can be consistently implemented.
- On p. 8, the application states that, according to the developer, "the specifications were changed to provide clarity on the hospice required exclusion." The following page says the "developer attests that the specifications have not changed, and that additional reliability testing was conducted but is directionally the same." Presumably, the "not" should be struck from the second quote. The application describes how multiple types of data can be used, yet the "specification" dimension of reliability is not addressed in the application, and the reliability testing does not get at this. The fundamental question is whether the available rates truly represent whether the children have received the vaccines in question. Presumably this was assessed and found to be ok in an earlier application.
- No concerns identified. The developer attests that the specifications have not changed, additional reliability testing was conducted but results suggest it is directionally the same.
- The developer conducted new reliability testing for this resubmission. The developer conducted the same type of testing as they did in the previous submission with new data. In the previous submission there was no feedback from the Standing Committee. They voted to accept the 2012 reliability evaluation without further discussion as the 2012 and 2014 evaluation were directionally the same. Reliability testing was conducted using signal to noise ratio (beta-binomial model) with 2018-2020 data. This was calculated for each of the ten vaccines and stratified by commercial plans (N=391) and Medicaid plans (N=239). Reliability ranged from a low of 0.81 (VZV/MMR) to a high of 0.94 (IPV) among the commercial plans and from a low of 0.83 (VZV/MMR) to a high of 0.95 (Influenza) among the Medicaid plans. The average commercial plan reliability ranged from 0.81 (MMR/VZV) to 0.94 (IPV/Pneumococcal conjugate). The average Medicaid plan reliability ranged from 0.83 (MMR/VZV) to 0.95 (Influenza). The developer stated that the results suggest a high level of reliability for all ten of the vaccine measures.

2a2. Reliability - Testing

- No concerns.
- No concerns.
- New reliability testing shows strong reliability.
- No concerns.
- No concerns.
- No concerns.

- Based on the information provided, there do not appear to be any significant concerns about the reliability of the measure. The developer conducted reliability testing using the same method as in previous submissions and with updated data.
- Testing was conducted for each individual vaccine rather than the composite, which is how the measure is intended to be used. However, the strong results from the (appropriate) beta-binomial analysis for the individual vaccine rates means that the composite is also likely to be highly reliable.
- None noted.
- No concerns.

2b1. Validity Testing

- No concerns.
- No concerns.
- Validity testing was done for only 4 of the vaccine rates.
- Validity tests were not conducted for all vaccinations. The validity data is insufficient.
- No concerns.
- Staff noted some concerns.
- Based on the information provided, there are some concerns with the testing results. One concern is that statistical testing with a P value was not provided, which is important to determine the significance of the results. Another concern is that validity testing was not provided for six of the vaccine rates, including Hepatitis A, Hepatitis B, HiB, Influenza, IPV, and Pneumococcal Conjugate. This raises questions about the completeness and thoroughness of the validity testing. Additionally, the results were only stratified by payer and not by other demographic factors, which may limit the generalizability of the results.
- In addition to correlating rates for childhood vaccines with rates in the same plans for similar adolescent vaccines, it would have been useful to examine convergent validity by correlating rates for the ten childhood vaccines.
- New validity testing conducted at the Accountable-Entity Level of Dtap, MMR, Rotavirus, VZV and correlated with adolescent vaccines with positive correlations observed ranging from 0.41-0.79. Validity tests were not provided for Hepatitis A, Hepatitis B, HiB, Influenza, IPV, or Pneumococcal Conjugate. The measure is not risk-adjusted or stratified. Preliminary rating for validity is insufficient.
- The measure only uses one set of specifications for this measure. Potential threats to validity exist. Empirical validity testing was not conducted on the measure as specified. Face validity was not assessed.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)

- No comment.
- No significant risks
- Evidence on exclusions not provided but seem reasonable.

- The measure is not risk adjusted.
- Not risk adjusted. Exclusions are appropriate.
- No concerns.
- Incomplete.
- No concerns. Risk adjustment N/A.
- The measure is not risk-adjusted or stratified. Preliminary rating for validity is insufficient.
- The measure is not risk-adjusted or stratified.

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)

- No comment.
- Validity is adequate.
- The ICR's indicate significant but not very large differences. Worth discussion.
- No.
- No.
- No concerns.
- Incomplete.
- No concerns.
- None noted.
- The developer attested that NCQA has an audit process to check to ensure the HEDIS measures are correctly identified and reported. Validity tests were not provided for Hepatitis A, Hepatitis B, HiB, Influenza, IPV, or Pneumococcal Conjugate.

3. Feasibility

- Registry data used. No concerns on data collection strategy.
- Feasibility quite evident.
- No concerns.
- No concerns about feasibility. This data has been collected and is operationalized.
- No concerns. Data elements are routinely generated during care delivery.
- Feasibility was demonstrated.
- While the data elements needed to compute the measure are routinely generated and used during care delivery, and are available in electronic sources, there may still be some challenges in putting the data collection strategy into operational use. These challenges may include consistency in coding and abstraction across different sources and practices, and the need for manual abstraction from paper records and registry data.

- The application describes the multiple types of data that can be used and anticipates simplification as electronic health records become more widespread. One imagines that by 2023, much of this has already occurred, and it would be useful to know the degree to which plans are not using electronic records to calculate these rates.
- The developer noted that this measure is collected through administrative data, electronic clinical data, paper records, and registry data to allow for widespread adoption across health plans and health care practices; there is an eCQM version of this measure. Field tests and HEDIS results continue to demonstrate that this measure is highly feasible and usable. Broad public use and dissemination of these measures is encouraged.
- The developer stated that the data elements needed to compute the measure are generated or collected by and used by healthcare personnel during the provision of care, coded by someone other than person obtaining original information, and abstracted from a record by someone other than the person obtaining original information. The developer attested that some data elements are in defined fields in electronic sources. This measure is collected through administrative data, electronic clinical data, paper records, and registry data to allow for widespread adoption across health plans and health care practices. The developer further stated that they anticipate that as electronic health records become more widespread, the reliance on paper record review will decrease. The developers also noted that there is an eCQM version of this measure. The developer stated that field tests and HEDIS results continue to demonstrate that this measure is highly feasible and usable.

4a. Use

- Widespread adoption of this measure nationwide across commercial and Medicaid.
- Accountability & transparency aligned with feedback.
- No concerns.
- Usability has increased over time.
- Data elements are routinely generated. Those measured have access to results and assistance to interpret. Feedback opportunity is present.
- Accountability and transparency were demonstrated. How feedback is used to improve clinical practice was not clear.
- The measure is publicly reported through four public reporting programs, including NCQA Health Plan Rating, NCQA Annual State of Health Care Quality, CMS Medicaid Child Core Set, and CMS Health Insurance Marketplaces – Quality Rating System. The performance results of the measure are disclosed and available outside of the organizations or practices whose performance is measured. The measure is currently being used for accountability purposes in two payment programs, CMS HER Incentive Program and Merit-Based Incentive Payment System (MIPS) Quality Payment Program, as well as a regulatory and accreditation program, NCQA Health Plan Accreditation, and a quality improvement with benchmarking program, Quality Compass. The information about the measure's use and performance results is accessible to the public, providing transparency and accountability for the organizations and practices being evaluated. Based on the information provided, it appears that those being measured, and other users have been given opportunities to provide feedback on the measure performance and implementation. The developer has publicly reported rates and created benchmarks to help plans understand their performance and has also provided technical

assistance through various platforms. Input is gathered through multi-stakeholder advisory panels, public comments, and clarifications through the Policy Clarification Support System. The developer also attested that questions and concerns were responded to and that changes were incorporated into the measure based on the feedback received.

- No concerns.
- Results of this measure are publicly reported (with benchmarks) and currently used in a current accountability program (HEDIS via Quality Compass tool and presented through the Policy Clarification Support System. No significant barriers to implementation have been reported.
- The developer attested that broad public use and dissemination of these measures is encouraged and that noncommercial uses do not require the consent of the measure developer. However, commercial use of a measure requires the prior written consent of NCQA. The measure is used for public reporting in NCQA Health Plan Rating, NCQA Annual State of Health Care Quality, CMS Medicaid Child Core Set, CMS Health Insurance Marketplaces – Quality Rating System and accountability programs, CMS HER Incentive Program, Merit-Based Incentive Payment System (MIPS) Quality Payment Program, Accreditation: NCQA Health Plan Accreditation, QI and Benchmarking: Quality Compass.

4a. Usability

- None.
- Use of administrative data facilitates the needed scale.
- My only question is at what point are overall performance and disparities for a measure good enough to consider making room for other measures for which performance is not as high?
- I do not see any unintended consequences.
- Developer reported "fluctuation" in performance rates attributed to Covid pandemic. This could be related to disruption in usual activities, including completing F2F healthcare appts required for immunizations, but could also be attributable to growing skepticism about immunizations accelerated by Covid vaccine misinformation. For this measure, fluctuations and disparities may have to account for new subgroups, including communities where large populations have opted against childhood vaccines.
- Benefits are clear, per ACIP.
- The performance results can be used to further the goal of high-quality, efficient healthcare by providing valuable insights into areas of improvement, identifying best practices, and responding to emerging challenges. It seems that the benefits of the measure outweigh any potential harms, as no negative consequences were identified during its implementation and the measure has received positive feedback from the Measure Applications Partnership.
- No concerns.
- No improvements, unexpected findings, potential harms were reported.
- Per the developer: the number of accountable entities included in the measure have increased since the last submission; performance rates generally remained high with some fluctuations which they stated may be attributable to the coronavirus 2019 (COVID-19) pandemic. To demonstrate this potential attribution, the developer cited a study that found that 43.5 percent

of patients within 2020 were not up to date on their childhood vaccines, The Measure Applications Partnership in 2013 recommended support for the measure to be added to Physician Compare/Value-Based Payment Modifier (VBPM) and for it to remain included in Physician Quality Reporting System (PQRS) as it was NQF-endorsed. The workgroup noted that the measure addresses a population not represented in the program measure set, promotes alignment across programs, settings, and public and private sector efforts, and addresses program goals/requirements.

5: Related and Competing Measures

- No additional steps needed.
- No concerns.
- There are no known competing measures.
- Appears harmonization has already been optimized.
- The influenza measure was noted. However, the application of this measure to children up to age 2 was not mentioned.
- Yes, there are related and competing measures for influenza immunization. The NQF #0041 Preventive Care and Screening: Influenza Immunization and NQF #1659 Influenza Immunization are related measures. The NQF #1407 Immunizations for Adolescents, NQF #3620 Adult Immunization Status, NQF #3484 Prenatal Immunization Status, NQF #0680 Percent of Residents Who Were Assessed and Appropriately Given the Seasonal Influenza Vaccine (Short-Stay), and NQF #0431 Influenza Vaccination Coverage Among Healthcare Personnel as related measures.
- No concerns.
- While there are competing measures, the developer attested that the measures are harmonized to the extent possible.
- Related Measures: NQF #0041 Preventive Care and Screening: Influenza Immunization, NQF #1659 Influenza Immunization. NQF staff identified the following additional measures as related measures: NQF #1407 Immunizations for Adolescents, NQF #3620 Adult Immunization Status, NQF #3484 Prenatal Immunization Status, NQF #0680 Percent of Residents Who Were Assessed and Appropriately Given the Seasonal Influenza Vaccine (Short-Stay), NQF #0431 Influenza Vaccination Coverage Among Healthcare Personnel Harmonization. The developer attested that the measures are harmonized to the extent possible.

Pre-evaluation Public and Member Comments

No comments were received.

NQF #1407 Immunizations for Adolescents (National Committee for Quality Assurance)

Pre-evaluation Standing Committee Comments

1a. Evidence

- The evidence for this measure is based on the Advisory Committee on Immunization Practices (ACIP) guidelines and is endorsed by the Centers for Disease Control and Prevention (CDC). The evidence supports the efficacy of these vaccines in increasing resistance to bacterial diseases and improving health, length, and quality of life. The evidence applies directly to the specific process being measured, which is the percentage of adolescents 13 years of age who received specific vaccines by their 13th birthday. The process of receiving specific vaccines relates to the desired outcome of improving health, length, and quality of life. The evidence supports this relationship through increased resistance to bacterial diseases. I am not aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission.
- Established with precedent.
- Agree that the evidence basis for the measure has not changed and that there is no need for repeated discussion and a vote on evidence.
- This is my main concern with this measure. In reviewing the developer's responses, the history from the 2018 cycle, and careful review of Algorithm 1 and Table 2 from the 2021 Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement, I have concerns. I appreciate the ACIP process, but it seems as though "no change in evidence" and "no change in guidance" is replacing data and a systematic review.
- This is a maintenance measure with evidence to support the measure focus.
- ACIP recommendations supports the measures focus. Should the COVID vaccine be added to this measure?
- Based on up-to-date ACIP evidence review & recommendation.
- Evidence for this measure remains unchanged since the last review, has moderate strength, and should be retained.
- Percentage of adolescents 13 years of age who had one dose of meningococcal conjugate vaccine, one tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccine, and have completed the human papillomavirus (HPV) vaccine series by their 13th birthday. This is a process measure. The developer attests that the underlying evidence for the measure has not changed since the last NQF endorsement review A Summary of prior review in 2018 was provided. The measure is based on evidence from the Advisory Committee on Immunization Practices (ACIP) guidelines, which are endorsed by the Centers for Disease Control and Prevention (CDC). Since the Committee's last full review, Human Papilloma Virus (HPV) vaccination has been added to this measure and additional evidence was provided for the HPV vaccine. The ACIP graded the recommendation as evidence type three which means observational studies or randomized controlled trials with notable limitations, though the limitations were not provided. The developer also noted that no new additions were made to the body of evidence related to the meningococcal and Tdap/Td vaccines since the previous submission. The Standing Committee agreed that this measure met the evidence criterion.

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

- Current performance data on the measure was provided by the developer. The data showed that the performance rates for both commercial and Medicaid health plans have a significant variability. This demonstrates a gap in care and variability in performance which warrant the need for a national performance measure. Disparities: The developer reported data on the

measure by population subgroups and provided evidence from the National Immunization Survey which showed disparities in immunization coverage among different racial and socioeconomic groups. However, no data was provided to support this claim.

- Gap & disparities evidence supports.
- HPV the only measure with a significant performance gap. Consider focus on just this measure? Disparities noted but not presented. Consider changing focus to disparities rather than mean performance?
- There is a gap in care that warrants a national performance measure. Would have liked more information about population subgroups and disparities (no data provided). Also, given increasing vaccine hesitancy across the country and in some concentrated areas/communities, it seems that geographic performance data might be useful (in the future).
- Disparity data is availability at the health plan level.
- Performance gap data were provided. There continue to be performance gaps.
- Medicaid vs. commercial could be regarded as evidence of disparities by socioeconomic position (SEP), but the developer does not explicitly make this case. Furthermore, in addition to the data provided, the developer should have provided summary data from the National Immunization Survey, which surely would have documented important disparities by race, ethnicity, SEP, etc.
- The measure can identify performance gaps for each vaccine, and stratification using demographic variables, such as race/ethnicity or socioeconomic status, is possible when the data are available to a plan. Disparities have been observed in uninsured, Black and Hispanic patients, and patients living below the federal poverty line, but no data was offered to document the differences.
- The developer reported data from the Healthcare Effectiveness Data and Information Set (HEDIS) reflecting the most recent 3 years, 2020, 2021, and 2022, of measurement for this measure. Performance data is summarized at the health plan level and summarized by mean, interquartile range, and standard deviation. Data is stratified by year and product line. There is a performance gap for Tdap, HPV, and Meningococcal for all three years, worse for HPV. Disparities: The developer states that the measure can be stratified by demographic variables, such as race/ethnicity or socioeconomic status, to assess the presence of health care disparities, if the data are available to a plan; cited evidence from the National Immunization Survey which they stated showed that national coverage with most routine childhood vaccines remained stable, however, disparities in immunization coverage have been seen in uninsured, Black and Hispanic patients, and patients living below the federal poverty line compared to individuals who were privately insured, White, or living at or above the federal poverty line. The developer did not provide data to support this claim.

2a1. Reliability - Specifications

- Based on the information provided, it seems that all data elements, codes with descriptors, and steps in the logic or calculation algorithm or other specifications are clearly defined and provided. There are no concerns about the likelihood that this measure can be consistently implemented. All necessary information is provided and there are no issues with data elements, codes or the algorithm.
- Reliability adequate for population studies.

- Measure specifications are clear and precise.
- Reliability specifications are clearly defined. No concerns.
- New reliability testing has been conducted with a preliminary finding of high reliability.
- No concerns.
- In section 2a, the application does not describe the measures specifications other than to say that they have not changed, and the "specification" dimension of reliability is not addressed in the application, and the reliability testing does not get at this. However, the feasibility section (3) indicates that multiple types of data can be used. The fundamental question is whether the available rates truly represent whether the adolescents have received the vaccines in question. Presumably this was assessed and found to be ok in an earlier application.
- No changes to measure specifications changed since the last review. New reliability analyses suggest high reliability for all 3 of the vaccine measures.
- Reliability testing conducted at the Accountable-Entity Level: The developer conducted the same type of testing as they did in the previous submission with new data. The Standing Committee noted that no data element reliability testing was completed though the measure uses multiple data sources. Reliability testing was conducted using signal to noise ratio (beta-binomial model) with 2018-2020 data. This was calculated for each of the 3 vaccines and stratified by commercial plans (N=391) and Medicaid plans (N=239). Reliability ranged from a low of 0.91 (HPV) to a high of 0.94 (Tdap/Meningococcal) among the commercial plans and from a low of 0.93 (Tdap) to a high of 0.95 (Meningococcal) among the Medicaid plans. The average commercial plan reliability ranged from 0.92 (HPV) to 0.94 (Meningococcal/Tdap). The average Medicaid plan reliability ranged from 0.93 (Tdap) to 0.95 (Meningococcal). The results suggest a high level of reliability for all 3 of the vaccine measures.

2a2. Reliability - Testing

- Based on the information provided, there is no concern about the reliability of the measure.
- No concerns.
- No concerns.
- No concerns.
- No concerns.
- No concerns.
- Testing was conducted for each individual vaccine rather than the composite, which is how the measure is intended to be used. However, the strong results from the (appropriate) beta-binomial analysis for the individual vaccine rates means that the composite is also likely to be highly reliable.
- None noted.
- No concerns.

2b1. Validity Testing

- Based on the information provided, there are no concerns with the testing results for validity.

The exclusions in the measure are consistent with the evidence and no patients or patient groups are inappropriately excluded. The measure is not risk adjusted, so there is no need for analysis of the relationship between potential social risk factor variables and the measure focus or the development and testing of a risk-adjustment strategy.

- No concerns.
- No, though P values not provided.
- No concerns.
- No concerns. New validity testing has been conducted.
- No concerns.
- In addition to correlating rates for adolescent vaccines with rates in the same plans for similar childhood vaccines, it would have been useful to examine convergent validity by correlating rates for the three adolescent vaccines.
- Validity testing was conducted at the Accountable-Entity Level suggests that the adolescent vaccine measure correlates positively with the childhood measure, but p-value statistics were not reported.
- This measure excludes patients who have a contraindication for the vaccine and patients who use hospice services during the measurement year. The developer did not conduct statistical analyses to determine the impact of the exclusions on the measure rates, noting low rates of reported plan exclusions.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)

- The exclusions in the measure are consistent with the evidence and no patients or patient groups are inappropriately excluded. The measure is not risk adjusted, so there is no need for analysis of the relationship between potential social risk factor variables and the measure focus or the development and testing of a risk-adjustment strategy.
- Evidence for consistent exclusions & adequate risk adjustment
- No concerns.
- Measure is not risk adjusted.
- The measure is not risk adjusted.
- No concerns.
- No concerns. Risk adjustment N/A.
- The measure is not risk adjusted. No meaningful differences identified. NCQA has an audit process to check to ensure the HEDIS measures are correctly identified and reported.
- The measure is not risk adjusted. The developer concluded there is no conceptual reason to risk-adjust a measure assessing vaccination rates.

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)

- The analyses indicate that the measure identifies meaningful differences about quality by

calculating an interquartile range and conducting independent sample t-tests between two randomly selected plans in each group. There is an audit process in place to handle missing data and ensure the accuracy of the HEDIS measures. The measure only uses one set of specifications, and thus there is no need for comparability analysis.

- Validity is adequate.
- No concerns, ICR's relatively small.
- Some data missing but not likely a threat to validity.
- No.
- No concerns.
- No concerns.
- The developer did not conduct statistical analyses to determine the impact of the exclusions on the measure rates but did provide information on prevalence of the exclusions by payer.
- The measure is not risk adjusted. The developer concluded there is no conceptual reason to risk-adjust a measure assessing vaccination rates. Meaningful Differences The developer calculated an interquartile range for each of the 3 vaccine rates and all 3 combined and then conducted an independent sample t-test between two randomly selected plans in each group (below 25th percentile and above 75th percentile). The IQR ranged from 9 to 13 percentage points for commercial plans and from 8 to 14 percentage points for Medicaid plans. The developer states all measure for both commercial and Medicaid plans were statistically different from zero using a p-value threshold of 0.05. Missing Data NCQA has an audit process to check to ensure the HEDIS measures are correctly identified and reported. If a data source is found to have missing data and the issue cannot be rectified, the auditor will assign a "materially biased" designation to the measure for the reporting plan and the rate will not be reported. Comparability The measure only uses one set of specifications for this measure.

3. Feasibility

- The data collection strategy is feasible as all required data elements are routinely generated and used during care delivery. There are no concerns about how the data collection strategy can be put into operational use.
- Feasibility quite evident.
- No concerns.
- No concerns regarding feasibility.
- No concerns about feasibility. This data has been collected and is operationalized.
- Feasibility was demonstrated.
- The application describes the multiple types of data that can be used but does not regard this as a feasibility issue. I am not so sure.
- Data elements needed to compute the measure are generated or collected by and used by healthcare personnel during the provision of care and are coded by someone other than the person obtaining original information from multiple data sources including administrative data,

electronic clinical data, paper records, and registry data.

- The developer reports that the data elements needed to compute the measure are generated or collected by and used by healthcare personnel during the provision of care and are coded by someone other than the person obtaining original information. The developer states that some data elements are in defined fields in electronic sources. The developer reports that to allow for widespread reporting across health plans and health care practices, if applicable, this measure is collected through multiple data sources including administrative data, electronic clinical data, paper records, and registry data. The developer noted that the measure is not currently developed as an eCQM. The developer states that the measure's current use and history of use demonstrates its feasibility for reporting entities. Data collection from medical records, administrative sources, and registry data is robust and well-supported.

4a. Use

- The measure is publicly reported through the NCQA Health Plan Rating program, NCQA Annual State of Health Care Quality Report, CMS Medicaid Child Core Set, CMS Health Insurance Marketplaces Quality Rating System, NCQA Health Plan Accreditation program, and the Quality Compass Program. The performance results are disclosed and available to the public through Consumer Reports, the NCQA website, and the NCQA Annual State of Health Care Quality Report. The measure is being used for various accountability applications including calculation of health plan ratings, better understanding of the quality of health care received by children enrolled in Medicaid and CHIP, scoring for accreditation of Medicare Advantage Health Plans, and as a tool for selecting a health plan, conducting competitor analysis, examining quality improvement, and benchmarking plan performance through the Quality Compass Program. Yes, those being measured have been given performance results and data, as well as assistance with interpreting the measure results and data. Those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation through NCQA's use of a consensus-based process to evaluate the measures, and by soliciting input through multiple methods such as vetting with multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. Feedback provided by those being measured and others has been considered when changes are incorporated into the measure through NCQA's regular evaluation process which includes input from multiple stakeholders.
- Accountability & transparency aligned with feedback.
- No concerns.
- Measure is being used widely, performance results are reported and publicly available. Also used for accreditation purposes.
- The data is publicly reported and used in an accountability program.
- Accountability and transparency were demonstrated.
- No concerns.
- The developer reports that the measure is publicly reported and currently in use in various accountability programs that include: NCQA Health Plan Rating program, ratings are based on performance on HEDIS measures nationally and regionally, for CMS Medicaid Child Core Set, CMS Health Insurance Marketplaces Quality Rating System, Medicare Advantage Health Plans and Quality Compass Program and creates benchmarks to help health plans understand how

they perform relative to others and published annually in the Quality Compass tool and presented at conferences and webinars. A consensus-based process is used regularly and includes input from multiple stakeholders.

- The developer reports that NCQA publicly reports rates across all plans and creates benchmarks to help health plans understand how they perform relative to others. The developer states that HEDIS results are published annually in the Quality Compass tool. Additionally, the developer presents data at various conferences and webinars and provides technical assistance through its Policy Clarification Support System. The developer reports that NCQA measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. Additionally, the developer states that several methods are used to solicit input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and a review of questions submitted to the Policy Clarification Support System.

4a. Usability

- Use of administrative data facilitates the needed scale.
- Performance on these measures seems to have leveled off over the past 3 years. Worth discussion.
- Wondering if performance results could be used to help identify communities where vaccine hesitancy is prevalent as additional processes (counseling) may be necessary to address vaccine hesitancy (among parents).
- I do not see any unintended consequences.
- How feedback is used to improve clinical practice was not clear.
- No concerns.
- The developer reports that the number of accountable entities has increased for this measure. Performance rates for this measure stayed high with some fluctuation. No identified unintended findings, risks, benefits, or harms were identified since implementation.
- The developer reports that the number of accountable entities has increased for this measure. Previous submission data showed from 2012-2014 an average of commercial plans reporting was 346 and for Medicaid it was 175 plans. Data submitted in 2019-2021 shows an average of 391 commercial plans and 240 Medicaid plans reporting. The developer reports that performance rates for this measure stayed high with some fluctuation, and that rate fluctuation may be a result of the COVID-19 pandemic. The developer states that there were no identified unintended findings for this measure during testing or since implementation. The developer states that there were no identified unexpected benefits for this measure during testing or since implementation. The measure was reviewed by the Measure Applications Partnership in 2013. The Workgroup recommended support for the measure to be added to Physician Compare/Value-Based Payment Modifier (VBPM) and for it to remain included in Physician Quality Reporting System (PQRS) as it was NQF-endorsed."

5: Related and Competing Measures

- No additional steps needed.
- No concerns.

- No competing measures. Not clear that this measure requires harmonization with the related measures identified by staff.
- There are no known competing measures.
- None.
- No concerns.
- No NQF-endorsed or non-NQF endorsed related or competing measures.
- The developer does not identify any NQF-endorsed or non-NQF endorsed related or competing measures. NQF staff identified the following related measures: NQF #0038 Childhood Immunization Status (CIS); NQF #3620 Adult Immunization Status; NQF #3484 Prenatal Immunization Status

Pre-evaluation Public and Member Comments

No comments were received.