



Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 3749e

Corresponding Measures:

Measure Title: Diagnostic Delay of Venous Thromboembolism (DOVE) in Primary Care

Measure Steward: Brigham and Women's Hospital

sp.02. Brief Description of Measure: This eCQM assesses the rate of delayed diagnosis of VTE in adults aged 18 years and older in the primary care setting. Delayed diagnosis is defined as a diagnosis of VTE that occurs >24 hours following the index primary care visit where symptoms for the VTE were first present (within 30 days). The target population for this measure is all patients, 18 years and older, across all payers.

1b.01. Developer Rationale:

sp.12. Numerator Statement: The subset of the denominator where the patient's VTE diagnosis occurs greater than 24 hours following a primary care visit (within 30 days).

sp.14. Denominator Statement:

All adult patients (age 18 years and older) presenting in primary care with VTE-related symptoms (see **table 1**), who are subsequently diagnosed with VTE following a primary care visit (within 30 days). VTE-related symptoms are identified in the EHR either as structured data (using the VTE-related symptoms value set, OID 2.16.840.1.113762.1.4.1206.51) or identified in unstructured data I clinical notes by a natural language processing (NLP) algorithm. A VTE diagnosis is defined using ICD billing codes, imaging codes, and RxNorm codes for therapeutic anticoagulants. All three codes must be present for an eligible VTE encounter.

cough	hypotension	lightheadedness
syncope	tachycardia	hemoptysis
shortness of breath	calf pain	leg pain
foot pain	Calf numbness	leg numbness
foot numbness	calf tingling	leg tingling
foot tingling	calf redness	leg redness
foot redness	calf swelling	leg swelling
foot swelling	calf tenderness	leg tenderness
foot tenderness	calf warmth	leg warmth

cough	hypotension	lightheadedness
foot warmth		

Table 1: VTE-related symptoms:

sp.16. Denominator Exclusions: This eCQM excludes patients who are receiving hospice or palliative care.

Measure Type: Outcome

sp.28. Data Source:

Electronic Health Records

sp.07. Level of Analysis:

Clinician: Group/Practice

Integrated Delivery System

IF Endorsement Maintenance – Original Endorsement Date:

Most Recent Endorsement Date:

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

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

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

1. Importance to Measure and Report

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

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

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]

[Response Ends]

1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

[Response Begins]

[Response Ends]

1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

2. Scientific Acceptability of Measure Properties

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

sp.01. Provide the measure title.

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

[Response Begins]

Diagnostic Delay of Venous Thromboembolism (DOVE) in Primary Care

[Response Ends]

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

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

[Response Begins]

This eCQM assesses the rate of delayed diagnosis of VTE in adults aged 18 years and older in the primary care setting. Delayed diagnosis is defined as a diagnosis of VTE that occurs >24 hours following the index primary care visit where symptoms for the VTE were first present (within 30 days). The target population for this measure is all patients, 18 years and older, across all payers.

[Response Ends]

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

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

Please do not select:

- Surgery: General

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

venous thromboembolism, primary care

[Response Ends]

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

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

delayed diagnosis

Safety

[Response Ends]

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

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

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

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Adults (Age >= 18)

[Response Ends]

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

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

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

Please do not select:

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

[Response Begins]

Clinician: Group/Practice

Integrated Delivery System

[Response Ends]

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

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

[Response Begins]

Outpatient Services

[Response Ends]

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

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

[Response Begins]

None available

[Response Ends]

sp.10. Indicate whether Health Quality Measure Format (HQMF) specifications are attached.

Attach the zipped output from the eQCM 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).

[Response Begins]

HQMF specifications are attached.

[Response Ends]

Attachment: 3749e_DOVE Human Readable 010423.zip

sp.11. Attach the simulated testing attachment.

All eQCMs require a simulated testing attachment to confirm that the HTML output from Bonnie testing (or testing of some other simulated data set) includes 100% coverage of measured patient population testing, with pass/fail test cases for each sub-population. This can be submitted in the form of a screenshot.

[Response Begins]

Testing is attached

[Response Ends]

Attachment: 3749e_DOVE Bonnie Testing.zip

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

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

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 3749e_Dove_Prjct_Data_Dictionary_01-03-2023.xlsx

Attachment: 3749e_Value Sets_(1).zip

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.13. State the numerator.

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

DO NOT include the rationale for the measure.

[Response Begins]

The subset of the denominator where the patient's VTE diagnosis occurs greater than 24 hours following a primary care visit (within 30 days).

[Response Ends]

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

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

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

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

[Response Begins]

A patient is included in the numerator if they are included in the denominator population and their VTE diagnosis occurs >24 hours following their primary care visit (within 30 days).

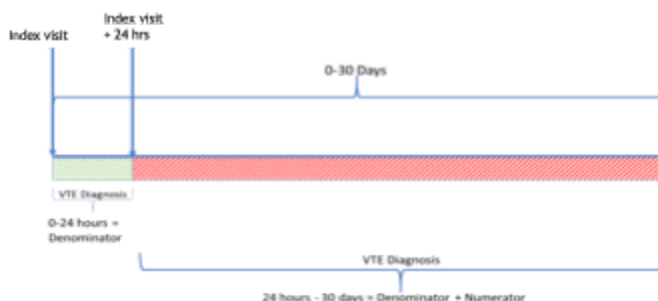


Figure 1: DOVE Numerator Inclusion Timeline

[Response Ends]

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

All adult patients (age 18 years and older) presenting in primary care with VTE-related symptoms (see **table 1**), who are subsequently diagnosed with VTE following a primary care visit (within 30 days). VTE-related symptoms are identified in the EHR either as structured data (using the VTE-related symptoms value set, OID 2.16.840.1.113762.1.4.1206.51) or identified in unstructured data I clinical notes by a natural language processing (NLP) algorithm. A VTE diagnosis is defined using ICD billing codes, imaging codes, and RxNorm codes for therapeutic anticoagulants. All three codes must be present for an eligible VTE encounter.

cough	hypotension	lightheadedness
syncope	tachycardia	hemoptysis
shortness of breath	calf pain	leg pain
foot pain	Calf numbness	leg numbness
foot numbness	calf tingling	leg tingling
foot tingling	calf redness	leg redness
foot redness	calf swelling	leg swelling
foot swelling	calf tenderness	leg tenderness
foot tenderness	calf warmth	leg warmth
foot warmth		

Table 1: VTE-related symptoms:

[Response Ends]

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

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

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

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

[Response Begins]

To be included in the measure cohort for analysis, patients must meet the following inclusion criteria.

1. Aged 18 years or older on the date of the primary care visit
2. All PCP visits in this measure must be performed by a provider with the following specialties: Nurse Practitioner (occupation), Physician (occupation), Medical practitioner (occupation), Technical healthcare occupation (occupation), Family medicine specialist (occupation), General practitioner assistant (occupation), General practitioner principal (occupation), Associate general practitioner (occupation)
3. Receive a diagnosis of Venous Thromboembolism within 30 days of their primary care visit. For a patient to have a VTE diagnosis, they must have all of the following VTE-related codes within the same encounter (see attached value sets for relevant codes): ICD-10 CM code for VTE, CPT codes for an imaging scan for VTE linked to the same encounter as the ICD-10 CM code, RxNorm order for therapeutic anticoagulants placed in the same encounter as the imaging scan.
4. Have no eligible VTE events within 6 months of the qualifying VTE event

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

This eCQM excludes patients who are receiving hospice or palliative care.

[Response Ends]

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

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

[Response Begins]

Patients who are on hospice or palliative care within 90 days of the eligible VTE encounter are excluded from this measure. Rationale: These patients have different care goals than non-hospice or palliative care which may affect their VTE diagnosis. Value Set: Hospice/Palliative Care (**table 2**).

Value Set Name	Steward	OID Number
Hospice care	Brigham and Women's Hospital	2.16.840.1.113762.1.4.1108.15
Palliative care	Brigham and Women's Hospital	2.16.840.1.113883.3.464.1003.101.12.1090

Table 2: Value Sets for Measure Exclusion Criteria

[Response Ends]

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

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

[Response Begins]

N/A, measure not stratified

[Response Ends]

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

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

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

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

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

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

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

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

[Response Begins]

Better quality = Lower score

[Response Ends]

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

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

[Response Begins]

Step 1: Define the target population:

Identify all patients aged 18 years or older who presented in primary care with VTE-related symptoms (identified in the clinical notes by the NLP algorithm) who are diagnosed with VTE following a primary care visit within 30 days.

Step 2: Define the denominator:

Identify qualifying VTE events. A qualifying VTE event is defined using the following three criteria within 30 days of the primary care visit (all must be present for measure inclusion):

- ICD-10 CM code for VTE.
- CPT imaging codes for VTE linked to the same encounter as the ICD billing codes relating to VTE.
- RxNorm anticoagulant order placed within the same encounter as the imaging scan.

Apply the exclusion criteria (patients in hospice or palliative care within 90 days of the eligible encounter) to all patients from the target population. If a VTE patient has more than one eligible presentation to primary care within 6 months after applying the exclusion criteria, we will use the first eligible admission and exclude subsequent eligible admissions in that period.

Step 3: Define the numerator:

Identify all patients from the denominator who had a VTE diagnosed >24 hours following a primary care visit (within 30 days) where the patient presented with VTE symptoms.

Step 4: Calculate the rate:

Divide the number of patients in the numerator (Step 3) by the number of patients in the denominator (Step 2) and multiply by 100. The measure is reported as a percentage: XX out of 100.

[Response Ends]

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

Examples of samples used for testing:

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

[Response Begins]

N/A, the measure is not based on a sample.

[Response Ends]

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

[Response Begins]

Electronic Health Records

[Response Ends]

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

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

[Response Begins]

At the time of Intent to Submit submission, this eCQM has been tested in two geographically different healthcare systems in Massachusetts (Site 1) and Kentucky (Site 2). Both sites are academic medical centers. Site 1 is a large, urban healthcare system with an interoperable electronic health record (EHR) system, meaning that EHR information can be shared across practices within the healthcare system. Site 2 is a smaller, semi-rural healthcare system with a non-interoperable EHR, meaning that EHR data cannot be transferred across practices within the healthcare system. A third testing site in Pennsylvania has been proposed for final measure testing and full NQF submission.

Site 1 uses EPIC EHR, data were collected from 2016-2020. Historical data from Site 2 were used from 2016-2020; the rationale for using historical data was to assess the eCQM performance in the Allscripts EHR system, prior to Site 2's transition to EPIC in 2020.

This eCQM leverages structured and unstructured data in the EHR. Structured data are used to assess patient demographic characteristics, inclusion/exclusion criteria, confirm VTE diagnoses, and measure the time between the initial primary care encounter and VTE finalizing date. Unstructured data from the clinical notes are used to identify VTE symptoms during the index primary care visit (when not present as structured data) and are computed into binary SNOMED codes for inclusion in the eCQM. Analysis of unstructured EHR data in clinical notes via natural language processing (NLP) algorithms has seen increasing use in clinical fields like radiology (Steinkamp

& Cook, 2021), oncology (Zeng et al., 2021; Savova et al., 2019), and post-operative VTE detection (Shi et al., 2021). VTE symptoms are also available as structured data using the VTE-related symptoms value set (OID 2.16.840.1.113762.1.4.1206.51).

References:

1. Steinkamp, J. and Cook, T.S., 2021. Basic Artificial Intelligence Techniques: Natural Language Processing of Radiology Reports. Radiologic Clinics, 59(6), pp.919-931.
2. Zeng, J., Banerjee, I., Henry, A.S., Wood, D.J., Shachter, R.D., Gensheimer, M.F. and Rubin, D.L., 2021. Natural language processing to identify cancer treatments with electronic medical records. JCO Clinical Cancer Informatics, 5, pp.379-393.
3. Savova, G.K., Danciu, I., Alamudun, F., Miller, T., Lin, C., Bitterman, D.S., Tourassi, G. and Warner, J.L., 2019. Use of Natural Language Processing to Extract Clinical Cancer Phenotypes from Electronic Medical Records Natural Language Processing for Cancer Phenotypes from EMRs. Cancer research, 79(21), pp.5463-5470.
4. Shi J, Hurdle JF, Johnson SA, Ferraro JP, Skarda DE, Finlayson SR, Samore MH, Bucher BT. Natural language processing for the surveillance of postoperative venous thromboembolism. Surgery. 2021 Oct 1;170(4):1175-82.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

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

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

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

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

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

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

AND

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

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

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

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

OR

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

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

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

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

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

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

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

Definitions

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

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity

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

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

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

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

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of \$25 in cost for an episode of care (e.g., \$5,000 v. \$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

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

[Response Begins]

Electronic Health Records

[Response Ends]

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

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

[Response Begins]

N/A, existing dataset not used. For measure testing, data were pulled from two U.S. academic healthcare systems using Epic and Allscripts EHRs. A third testing site will be included in the final measure submission.

[Response Ends]

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

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

[Response Begins]

Site 1: 01/06/2016 - 12/31/2021

Site 2: 12/01/2016 - 12/31/2020

Third site: to be determined.

[Response Ends]

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

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

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

Please do not select:

- Clinician: Clinician
- Population: Population

[Response Begins]

Clinician: Group/Practice

Integrated Delivery System

[Response Ends]

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

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

[Response Begins]

The Site 1 sample included a total of 220 primary care sites. To promote clinical meaningfulness, the largest 50 clinician groups from Site 1 were assessed in measure testing. The 50 largest clinician groups in the Site 1 sample represented a total of 3,072 encounters (ranging from 11-1,035) that met the measure inclusion criteria.

As a non-interoperable and semi-rural site, Site 2 technical experts faced difficulties in accurately capturing clinician group levels, and this site was assessed as a single clinician group at the facility-level. This is noted as a limitation of testing. The Site 2 sample represented a total of 245 encounters that met the measure inclusion criteria. As a semi-rural, non-interoperable healthcare system, a larger proportion of encounters in Site 2 did not meet the inclusion criteria of having a primary care encounter and subsequent VTE diagnosis within the same healthcare system compared to Site 1 (61.23% of encounters did not meet inclusion criteria, compared to 35.14%). Accessing care across sites is a limitation of eQMs in non-interoperable systems and is not limited to the DOVE eQM. Based on testing in Site 2, we have determined that the measure would be most meaningful when used within an integrated care delivery network and we are therefore adding a third testing site that is an integrated care delivery network.

[Response Ends]

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

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

[Response Begins]

Table 3 displays the descriptive statistics of patients who met the inclusion criteria for the DOVE eCQM. **Table 4** displays the descriptive statistics of patients who did not meet the inclusion criteria for the DOVE eCQM. There is no minimum sample size requirement for this measure. Due to data sharing limitations, standard deviations for mean age, number of VTE symptoms, and income level were not calculated for the included and excluded samples in Site 2.

Variable	Site 1	Site 2
Number of total encounters	5,510	632
Encounters included in the measure (%)	3,574 (64.86)	245 (38.77)
Encounters excluded from the measure (%)	1,936 (35.14)	387 (61.23)
Number of delayed VTE diagnosis events	2,583	189
Site delayed diagnosis rate	72.27%	77.14%
Number of clinician groups	220	1
Age:		
Mean age at VTE (SD)	65.93 (15.14)	58.14 (N/A)
Age ≥65 (%)	2,079 (58.17)	84 (34.29)
Age <65 (%)	1,495 (41.83)	161 (65.71)
Self-Reported Race (%):		
Black/African American (%)	312 (8.73)	24 (9.80)
White (%)	2,934 (82.09)	221 (90.20)
Other* (%)	328 (9.18)	0 (0)
Self-Reported Ethnicity (%):		
Hispanic	230 (6.44)	4 (1.63)
Non-Hispanic	3,275 (91.63)	236 (96.33)
Missing ethnicity	69 (1.93)	5 (2.04)
Insurance Type (%):		
Public Insurance	1,750 (48.96)	162 (66.12)
Private Insurance	1,436 (40.18)	83 (33.88)
Missing insurance	388 (10.86)	0 (0)
English as a first language (%)	3,301 (92.36)	241 (98.37)
Female (%)	1,831 (51.23)	129 (52.65)
Mean number of VTE symptoms (SD)	2.31 (1.34)	1.4 (N/A)
Median income (via ZIP Code) (SD)	\$75,000 (\$27,376)	\$38,254 (N/A)
<i>*Other category includes Asian, American Indian or Alaska Native, and race self-reported as "other"</i>		

Table 3: Descriptive Statistics of the Included Sample

Variable	Site 1	Site 2
Number of total encounters	5,510	632
Encounters excluded from the measure (%)	1,936 (35.14)	387 (61.23)
Number of clinician groups	175	1
Age:		
Mean age at VTE (SD)	64.10 (15.33)	55.61 (N/A)
Age ≥65 (%)	1,042 (53.82)	131 (33.85)
Age <65 (%)	894 (46.18)	256 (66.15)
Self-Reported Race (%):		
Black/African American (%)	178 (9.19)	58 (14.99)
White (%)	1,586 (81.92)	326 (84.24)
Other* (%)	172 (8.88)	3 (0.78)
Self-Reported Ethnicity (%):		
Hispanic	139 (7.18)	9 (2.33)
Non-Hispanic	1,757 (90.75)	373 (96.38)
Missing ethnicity	40 (2.07)	5 (1.29)
Insurance Type (%):		
Public Insurance	900 (46.49)	298 (77.0)
Private Insurance	816 (42.15)	85 (21.96)
Missing insurance	220 (11.36)	4 (1.03)
English as a first language (%)	1,769 (91.37)	375 (96.90)
Female (%)	933 (48.19)	199 (51.42)
Median income (via ZIP Code) (SD)	\$73,770 (\$26,608)	\$38,965 (N/A)
<i>*Other category includes Asian, American Indian or Alaska Native, and race self-reported as "other"</i>		

Table 4: Descriptive Statistics of the Excluded Sample

[Response Ends]

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

[Response Begins]

N/A, all testing was conducted with the sample described above.

[Response Ends]

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

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

[Response Begins]

The following social risk factors were available and analyzed:

- Age
- Sex
- Self-reported race
- Self-reported ethnicity
- ZIP Code (proxy for median income)*
- Insurance type (public, private)
- English as a first language

*via U.S. Census Bureau American Community Survey

[Response Ends]

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

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

Choose one or both levels.

[Response Begins]

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

[Response Ends]

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

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

[Response Begins]

The following forms of reliability testing were conducted:

- NLP Algorithm Accuracy: used to extract VTE-related symptoms from the EHR clinical notes
- VTE Phenotyping Algorithm Accuracy: used to determine VTE diagnoses for denominator inclusion

NLP Phenotyping Algorithm Accuracy Testing:

We developed a rule-based symptom extractor to identify VTE symptoms in primary care clinical notes. We used a random sample of 329 patients from Site 1 who were diagnosed with VTE between 2016 and 2021 and had primary care visits within the 30 days prior to their VTE diagnosis. Batches of 10-15 patients were randomly selected for inclusion. We manually extracted notes from their visits and pasted them into a text file. We then split the notes into sentences using the Medical Text Extraction, Reasoning, and Mapping System (MTERMS) natural language processing system. We used a rule-based approach of regular expressions to identify terms from a lexicon derived from a set of VTE symptoms. Symptoms were reviewed and revised over the course of the study in accordance with physician expert guidance. We measured precision (PPV), recall (sensitivity), specificity, and NPV with a total of 26 rounds for patients with a VTE diagnosis, and 5 rounds for patients with no VTE diagnosis.

VTE Phenotyping Algorithm Accuracy

Phenotyping algorithm accuracy refers to the process we developed to define VTE events in the primary care setting using routinely available EHR data. VTE events are not always defined in the EHR, thus we developed and tested a phenotyping algorithm to accurately define and quantify VTEs.

Code Selection: Based on findings from a literature review conducted with the Harvard Countway Library and feedback from stakeholders and our technical expert panel (TEP), we determined that diagnosing a new VTE case should utilize the following three data elements in the EHR:

- ICD-10 CM billing codes
- CPT imaging codes
- RxNorm codes for therapeutic anticoagulant treatment

Chart Review Sample: Records for the target population of patients, those aged 18 years and older who had an ICD-10 CM code for VTE from December 2016 – January 2020 from Site 1 were extracted from the EHR using Clarity, EPIC's database. From this cohort, we selected those who had a primary care visit (defined as an office visit with an internal medicine, general medicine, or family medicine provider) within 30 days of their ICD-10 CM code being added. We then examined the patients who also had a VTE-related imaging code linked to the same encounter as the ICD-10 CM code and had an anticoagulant ordered or administered 6 hours prior to or following their imaging scan. To ensure that VTE events identified were new and not existing cases, we excluded patients who had VTE diagnoses within 6 months prior to the index VTE diagnosis date (defined as the “wash-out” period).

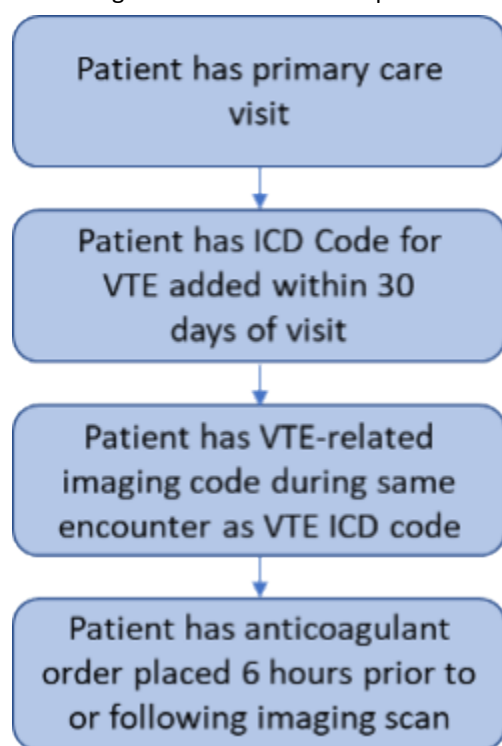


Figure 2: Novel VTE Phenotype cohort development and analytic pipeline

Chart Reviews and Algorithm Performance: we calculated the accuracy novel VTE phenotyping pipeline by calculating the positive predictive value (PPV), the negative predictive value (NPV), the sensitivity, and the specificity of chart reviews.

- PPV describes to the percentage of patients our algorithm indicates as having a positive VTE, who do have a positive VTE. Chart reviews were performed on 500 Site 1 patients who the algorithm defined as having a new VTE event (referred to as the VTE cohort).

- NPV describes the percentage of patients our algorithm indicates do not have VTE, who do not have a positive VTE.
- Sensitivity refers to the algorithm's ability to correctly classify an individual as "VTE-positive".
- Specificity refers to the ability to correctly classify an individual as "VTE-negative".

We performed chart reviews on distinct samples for each measurement (PPV, NPV, sensitivity, specificity). In chart reviews, the trained chart abstracter was deemed the "gold standard" to compare algorithm accuracy. The chart abstracter examined each of the patient's imaging results from the identified encounter to determine the presence or absence of a VTE as noted by the "imaging indication". Patients were considered to have a positive VTE diagnosis if the imaging scan noted the presence of a VTE.

A "true positive" was defined as a patient who was found to have a VTE during the encounter by both the algorithm and the chart abstracter. A "false positive" occurred if the algorithm identified a VTE during the encounter and the chart abstracter did not.

A "true negative" was defined as a patient who was not found to have a VTE during the encounter by both the algorithm and the chart abstracter. A "false negative" occurred if the algorithm did not identify a VTE during the encounter but the chart abstracter did. Using the true positive, true negative, false positive, and false negative rates, we calculated the algorithm's PPV, NPV, sensitivity, and specificity.

[Response Ends]

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

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

[Response Begins]

NLP Algorithm Accuracy Testing:

26 rounds of chart review were conducted with patients who had a VTE diagnosis. 5 rounds of chart reviews were conducted with patients without VTEs. Each round averaged 676 sentences of clinical notes.

Chart reviews were an iterative process, where the findings from one round would inform specification changes to be added to the next round. Annotators reviewed each round of chart review results and provided feedback for algorithm specifications:

- **Negation:** *patient does not report chest pain*
- **Context:** *concerns with leg pain were resolved*
- **Misspelling:** *patient reports leg swelling and couhg*
- **Search distance:** *swollen vein R medial ankle 3 weeks ago ... was very tender to touch*
- **Symptom attributed to wrong body part:** *worsening R hip pain as well as recent development of R leg, ankle and foot erythema*

	Round 1 (n=673)	Round 26 (n=938)
Precision (PPV)	0.50	1.00
Recall (sensitivity)	0.86	1.00
Specificity	0.93	1.00

	Round 1 (n=673)	Round 26 (n=938)
NPV	0.99	1.00

Table 5: NLP Algorithm Accuracy Testing for Patients with a VTE Diagnosis (Abridged)

	Round 1 (n=281)	Round 5 (n=912)
Precision (PPV)	0.53	0.850
Recall (sensitivity)	1.00	0.90
Specificity	0.974	1.00
NPV	1.00	1.00

Table 6: NLP Algorithm Accuracy Testing for Patients with no VTE Diagnosis (Abridged)

VTE Phenotyping Algorithm Accuracy

Code Selection Process: Following the stakeholder feedback and literature review with the Harvard Countway Librarian, we harmonized the ICD-10 CM code value set for VTE with an additional measure developed in 2021 by the Brigham and Women’s team entitled “Risk-Standardized major bleeding and venous thromboembolism rate following elective primary total hip arthroplasty and/or total knee arthroplasty electronic clinical quality measure”. Additional input from clinicians and healthcare experts on the TEP validated the imaging and RxNorm codes selected for this measure. The complete list of value sets used can be seen in **table 7** below.

Code System	OID	Description
ICD-10 CM	2.16.840.1.113762.1.4.1206.49	ICD codes used to code bill for a VTE-related service.
CPT	2.16.840.1.113762.1.4.1206.47	Imaging codes used to scan for a VTE.
RxNorm	2.16.840.1.113762.1.4.1206.19	RxNorm codes for medication used to treat a VTE.

Table 7: Value set codes used to indicate a VTE

PPV: The total “VTE cohort” for algorithm testing (patients who the algorithm identified as having a VTE event) consisted of 3,612 patients. Chart reviews were performed on a random sample of 500 of the 3,612 patients who fell into the “VTE cohort” as defined by our algorithm. Following chart review, 479/500 patients reviewed had a new, true diagnosis of VTE at the encounter determined by the chart abstractor using the diagnostic pipeline of ICD-10 CM codes, imaging codes, and RxNorm codes for anticoagulants. With 479 true positives and 21 false positives, our algorithm’s PPV was 95.80%.

Most of the false positives identified were instances where the provider suspected a pulmonary embolism (PE), and conducted imaging for PE, but instead found a pleural effusion, which was treated with anticoagulants. Therefore, the event was billed as a VTE, imaged as if it were a VTE but ruled out, and was treated like it were a VTE. As a result, our algorithm incorrectly noted these cases as a VTE.

NPV: Of the 500 randomly reviewed patients selected to determine the pipeline’s NPV, we found that no patients had a true VTE. Therefore, our algorithm correctly excluded all these patients producing 500 true negatives. Thus, using equation 2, this algorithm’s NPV is 100%.

Sensitivity and Specificity: Using the true positive, false positive, true negative, and false negative rates, our algorithm produced sensitivity and specificity rates of 100% and 95.69%, respectively.

Data element	Accuracy
PPV	95.80%
NPV	100%
Sensitivity	100%
Specificity	95.69%

Table 8: Chart Review Results

[Response Ends]

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

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

[Response Begins]

NLP Algorithm Accuracy Testing:

Our NLP algorithm can accurately define VTE-related symptoms from unstructured data in the EHR to identify patients with and without VTEs. Unstructured data has previously been inaccessible in eQMs, meaning that an estimated 80% of data in the EHR was inaccessible for quality measurement (De Boe, 2014; Martin-Sanchez & Verspoor, 2014). NLP technology in eQMs is particularly powerful for complicated disease conditions in large-scale patient populations, like VTE diagnosis in integrated healthcare systems.

VTE Phenotyping Algorithm Accuracy

The final VTE phenotyping algorithm was successful in accurately and reliably identifying VTE cases from structured data in ICD-10, imaging, and RxNorm codes.

In a chart review of 1,000 random patient encounters, our approach, which uses all three code types, was superior to previous approaches that used only one or two of these codes (**Figure 3**):

- DOVE eQM methods of ICD-10 CM, imaging, and RxNorm codes PPV = 95.8%
- ICD-10 only PPV = 64%
- ICD-10 and imaging (as done in other studies) PPV = 75%



Figure 3: Approaches to Identifying VTE in the EHR

Shi et al. (2021) developed a natural language processing (NLP) tool to detect postoperative venous thromboembolism from free-text EHR notes, similar to our approach. Internal validation demonstrated a sensitivity of 71% and specificity of 99%, compared to our sensitivity of 100% and specificity of 95.69%. In the two healthcare systems tested, this NLP approach demonstrated superior performance in DVT surveillance than existing tools, and similar performance in PE surveillance compared to existing tools. This study shows that NLP tools can effectively identify VTE events, and there is a need for more sensitive tools to identify VTE events using EHR notes in the primary care setting.

References:

1. De Boe, B., 2014. Use Cases for Unstructured Data - Intersystems White Paper, InterSystems Corporation. <http://www.odbms.org/wp-content/uploads/2014/08/Use-Cases-for-Unstructured-Data-White-Paper.pdf>

2. Martin-Sanchez, F. and Verspoor, K., 2014. Big data in medicine is driving big changes. *Yearbook of medical informatics*, 23(01), pp.14-20.
3. Shi J, Hurdle JF, Johnson SA, Ferraro JP, Skarda DE, Finlayson SR, Samore MH, Bucher BT. Natural language processing for the surveillance of postoperative venous thromboembolism. *Surgery*. 2021 Oct 1;170(4):1175-82.

[Response Ends]

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

[Response Begins]

Patient or Encounter-Level (data element validity must address ALL critical data elements)

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

[Response Ends]

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

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

[Response Begins]

Validity testing was assessed in the following ways:

- Data element availability
- Face validity
- Rate calculation

The measure development team plans to conduct public comment prior to full submission.

Data element availability

To confirm that all data elements needed to calculate the eCQM were commonly populated in the EHR, the eCQM development team analyzed the availability of the following variables from both sites:

- Age at admission
- Sex
- Race/ethnicity
- Insurance type
- Condition
- Primary care encounter
- VTE symptoms at the primary care encounter
- CPT imaging codes for VTE
- RxNorm therapeutic anticoagulant orders for VTE
- ICD billing codes related to VTE

Face validity

The objective of face validity testing was to demonstrate that this measure would be meaningful and beneficial to providers, patients, and informatics professionals, from the perspective of experts in the field. As a part of the validity testing process, we provided the TEP with several opportunities during the measure development process to suggest improvements/refinements to the measure to ensure optimal performance. The TEP consists of 6 members, three clinicians, one EHR expert, and two patient perspectives.

Rate Calculation

We calculated the DOVE eCQM rate in the largest 50 practices in Site 1, and at the facility level for Site 2. The purpose of this was to assess how the eCQM performed in a variety of settings and sizes of primary care facilities.

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

Data element availability

We validated that patient demographics were consistently present in the EHR (**Table 9**).

	Site 1 Available (%)	Site 1 Missing (%)	Site 2 Available (%)	Site 2 Missing (%)
Total eligible encounters:	3574	N/A	245	N/A
Age at VTE	3575 (100)	0 (0)	245 (100)	0 (0)
Sex	3576 (100)	0 (0)	245 (100)	0 (0)
Race	3577 (100)	0 (0)	245 (100)	0 (0)
Ethnicity	3518 (98.43)	56 (1.57)	240 (98.98)	5 (1.02)
Insurance type	3186 (89.14)	388 (10.86)	245 (100)	0 (0)
Language	3577 (100)	0 (0)	245 (100)	0 (0)
≥1 VTE symptom*	3578 (100)	0 (0)	245 (100)	0 (0)
Primary care encounter*	3579 (100)	0 (0)	245 (100)	0 (0)
VTE imaging scan*	3580 (100)	0 (0)	245 (100)	0 (0)
RxNorm anticoagulant order*	3581 (100)	0 (0)	245 (100)	0 (0)
VTE-related ICD billing codes*	3582 (100)	0 (0)	245 (100)	0 (0)
<i>*required for measure calculation</i>				

Table 9: Frequency of Data Element Availability

The only variable with >10% missing data was Site 1 insurance type. As this measure is not risk-adjusted and includes all payers, missing insurance information is not expected to impact measure calculation, and this systematic assessment or data imputation has not been conducted at this time.

Face validity

In the most recent TEP meeting (July 2022), TEP members were asked if they agreed with the following statement about the DOVE eCQM: **“The VTE Diagnostic Delay in Primary Care eCQM, as specified, can be used to distinguish good from poor clinician group-level quality related to patient safety.”** The final vote was 5/5 in agreement with the voting statement among present members. 1 member was absent and did not vote.

#3749e Diagnostic Delay of Venous Thromboembolism (DOVE) in Primary Care, Submission Last
Updated: Jan 09, 2023

Rate Calculation

Across the largest 50 primary care sites, DOVE rate in Site 1 was 71.26% (SD=0.12) with group-level rates ranging from 41.67%-92.31%. The overall DOVE rate in Site 1 across all 220 primary care sites was 72.27 % (SD=0.36). The facility-level DOVE rate in Site 2 was 77.14%.

Primary Care Site Identifier (organized by size)	Denominator Encounters	Numerator Encounters	Delayed VTE Diagnosis Rate
1	1,035	745	71.98%
2	473	373	78.86%
3	118	91	77.12%
4	115	95	82.61%
5	88	44	50.00%
6	65	46	70.77%
7	60	47	78.33%
8	53	41	77.36%
9	60	35	58.33%
10	44	33	75.00%
11	45	32	71.11%
12	39	32	82.05%
13	39	30	76.92%
14	38	26	68.42%
15	37	27	72.97%
16	38	20	52.63%
17	35	31	88.57%
18	37	27	72.97%
19	35	27	77.14%
20	31	20	64.52%
21	29	17	58.62%
22	29	25	86.21%
23	28	21	75.00%
24	27	20	74.07%
25	25	19	76.00%
26	25	16	64.00%
27	26	12	46.15%
28	28	19	67.86%
29	22	13	59.09%
30	24	16	66.67%
31	22	15	68.18%
32	22	14	63.64%
33	22	13	59.09%

Primary Care Site Identifier (organized by size)	Denominator Encounters	Numerator Encounters	Delayed VTE Diagnosis Rate
34	20	14	70.00%
35	17	12	70.59%
36	19	12	63.16%
37	16	12	75.00%
38	17	8	47.06%
39	17	12	70.59%
40	15	12	80.00%
41	15	7	46.67%
42	16	9	56.25%
43	14	12	85.71%
44	14	9	64.29%
45	14	6	42.86%
46	16	10	62.50%
47	13	12	92.31%*
48	12	5	41.67%**
49	12	7	58.33%
50	11	10	90.91%
Site 2 (facility level)	245	189	77.14%
<i>*Highest group level DOVE rate</i>			
<i>** Lowest group level DOVE rate</i>			

Table 10: Rate Calculation For Site 1 (Largest 50 practices) and Site 2 (Facility-level)

[Response Ends]

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

[Response Begins]

Data element availability

All data elements needed to calculate the eCQM are commonly available in the EHR. Some demographic variables (e.g., ethnicity and insurance status) demonstrated mixed availability, these variables do not impact measure calculation as they are not required for the inclusion, numerator, or denominator criteria, and this measure does not utilize a risk adjustment model that incorporates social risk variables.

Face validity

Face validity was established by a panel of experts who agreed that the measure is an accurate reflection of quality and that it can be used to distinguish between good and poor quality.

Rate Calculation

The measure score can be interpreted to reflect the quality of care provided to patients, and adequately identifies differences in quality as seen by the variation in rates across the largest 50 sites in Site 1, and facility level in Site 2. We identified consistently high DOVE rates across group practice sizes and healthcare systems, which demonstrates an opportunity for routine measurement and quality improvement.

Diagnosis of VTE is difficult due to the variability of documentation in EHR notes (Pellathy et al., 2021). Earlier diagnoses of VTE may reduce the morbidity and mortality associated with the dangerous condition (Dalen et al., 2002; Ozsu et al., 2011), meaning that more proximal diagnoses can promote patient safety. The lack of a standard definition of VTE, as well as the low performance of existing identification algorithms points to a need for the novel, data-driven DOVE eQIM. Measuring and reporting delayed VTE diagnosis rates will inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve the quality of care received by patients.

Site 2 was assessed at the facility level. As this measure includes an index primary care visit and a VTE event within 30 days, patients who receive their primary care visit in the testing sites but travel elsewhere for VTE treatment cannot be included in the measure. This was a limitation of our second testing site, which is not an integrated care network and does not have an interoperable EHR across sites, thus many patients were excluded from the measure, and clinician group levels could not be calculated. Although this measure is intended for use in integrated delivery networks, the high DOVE rate in Site 2 demonstrates that this measure can be used for quality improvement purposes in non-integrated healthcare systems.

References:

1. Pellathy T, Saul M, Clermont G, Dubrawski AW, Pinsky MR, Hravnak M. Accuracy of identifying hospital acquired venous thromboembolism by administrative coding: implications for big data and machine learning research. *Journal of Clinical Monitoring and Computing*. 2021 Feb 8:1-9.
2. Dalen JE. Pulmonary embolism: what have we learned since Virchow?: natural history, pathophysiology, and diagnosis. *Chest*. 2002 Oct 1;122(4):1440-56.
3. Ozsu S, Oztuna F, Bulbul Y, Topbas M, Ozlu T, Kosucu P, Ozsu A. The role of risk factors in delayed diagnosis of pulmonary embolism. *The American journal of emergency medicine*. 2011 Jan 1;29(1):26-32.

[Response Ends]

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

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

[Response Begins]

Currently, there are no federal-level measurement tools in place to track VTE events, or delayed diagnosis of VTE events, so the DOVE rates identified in two geographically different U.S. healthcare systems cannot be compared against an existing metric tool.

To assess clinically and practically meaningful differences in performance measure scores among our samples, we stratified clinician groups by encounter sample sizes into five cohorts and assessed the overall DOVE rate and range. The goal of this subgroup analysis was to understand if this measure can be meaningful in primary care clinician groups in both larger and smaller practices.

Clinician groups from Site 1 were stratified into four cohorts: >100 encounters during the study period, 50-99 encounters, 25-49 encounters, and <25 encounters. Due to limitations in group-level analysis in Site 2, Site 2 was assessed at the facility level.

[Response Ends]

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

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

[Response Begins]

In the subgroup analysis, rates between cohorts ranged from 64.90%-77.14%, and variation within cohorts is seen in each cohort rate range (**Table 11**).

Cohort	Cohort sample size	# Of groups in the cohort	Total denominator count	Total numerator count	Cohort rate (%) (SD)	Rate range (%)
Site 1A	>100	4	1,649	1,239	75.14 (0.04)	71.98-82.61
Site 1B	50-99	5	307	200	65.15 (0.12)	50.00-78.33
Site 1C	25-49	19	619	440	71.08 (0.11)	46.15-88.57
Site 1D	<25	22	359	233	64.90 (0.14)	41.67-92.31
Site 2	Facility level	N/A	245	189	77.14 (N/A)	N/A

Table 11: Sub-analysis by clinician group sample size

[Response Ends]

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

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

[Response Begins]

In the subgroup analysis, high rates with some variation were identified in each cohort. This demonstrates that the DOVE eQIM may be clinically and practically meaningful for understanding delayed diagnosis rates across sizes of clinician groups and can be used by clinician groups regardless of the practice size. The variation in rates points to opportunities for quality improvement at the clinician group level.

[Response Ends]

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

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

[Response Begins]

Patients cannot be included in this measure without the presence of ICD-10 billing codes related to VTE, CPT imaging codes for a VTE scan, and RxNorm codes for therapeutic anticoagulants, which together are used to identify the eligible VTE event. ICD-10 billing codes, CPT imaging codes, and RxNorm medication codes are directly

tied to care functions, are used for billing, and are less likely to be missing than other structured EHR data, like demographic information. Additionally, this measure does not rely on demographic variables for risk adjustment, further limiting the potential for missing data to impact the validity of this measure. There are no data types in this measure that rely on patient response, non-response is not a concern in this measure.

As part of validity testing, we assessed the frequency of data elements needed to calculate the measure (ICD billing codes related to VTE, RxNorm codes for therapeutic anticoagulants, and CPT imaging codes related to VTE), as well as the availability of demographic data. During algorithm development, we also assessed the negative predictive value to ensure that all cases that the algorithm defined as non-VTE events were in fact non-VTE events.

[Response Ends]

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

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

[Response Begins]

As seen in **Table 12**, all data elements required for encounter inclusion in the measure and measure calculation are commonly available within the EHR. In Site 1, some missing data was seen in ethnicity and insurance type, this will not impact measure calculation as this eCQM is not risk-adjusted.

During algorithm development, the measure's NPV was 100%, meaning all cases that the algorithm assigned as negative were true negative cases following trained chart review.

	Site 1 Available (%)	Site 1 Missing (%)	Site 2 Available (%)	Site 2 Missing (%)
Total eligible encounters:	3574 (100)	n/a	245	N/A
Age at VTE	3575 (100)	0 (0)	245 (100)	0 (0)
Sex	3576 (100)	0 (0)	245 (100)	0 (0)
Race	3577 (100)	0 (0)	245 (100)	0 (0)
Ethnicity	3518 (98.43)	56 (1.57)	240 (98.98)	5 (1.02)
Insurance type	3186 (89.14)	388 (10.86)	245 (100)	0 (0)
Language	3577 (100)	0 (0)	245 (100)	0 (0)
≥1 VTE symptom*	3578 (100)	0 (0)	245 (100)	0 (0)
Primary care encounter*	3579 (100)	0 (0)	245 (100)	0 (0)
VTE imaging scan*	3580 (100)	0 (0)	245 (100)	0 (0)
RxNorm anticoagulant order*	3581 (100)	0 (0)	245 (100)	0 (0)
VTE-related ICD billing codes*	3582 (100)	0 (0)	245 (100)	0 (0)
<i>*required for measure calculation</i>				

Table 12: Frequency of Data Elements

[Response Ends]

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

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

[Response Begins]

This review demonstrated that the data elements needed to calculate this measure are routinely available within the EHR, and the VTE algorithm can reliably distinguish between VTE and non-VTE events. Missing data is not expected to bias the results of this measure.

We did find through testing in a large nonintegrated care delivery network (most patients rural and nonmetro) where many patients get their follow-up care out of network, leading to a loss of patient encounters (the rate of delayed VTE diagnosis is still extremely high, but the number of encounters in the numerator is low). Therefore, we recommend that the measure is implemented in an integrated care delivery network.

[Response Ends]

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

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

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

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

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

[Response Begins]

This eCQM excludes patients who are in hospice or palliative care within 6 months of an otherwise eligible VTE event. These exclusions were supported by our technical expert panel comprised of clinicians, EHR experts, and patient representatives. Given the low frequency of VTE events and lower frequency of VTEs comorbid with hospice and palliative care, the impact of these exclusions is expected to be minimal. To ensure a low frequency of hospice and palliative care-related exclusions, we assessed the frequency of the exclusion on the target population (adults diagnosed with VTE events) in both sites.

[Response Ends]

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

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

[Response Begins]

In the Site 1 sample, 37 otherwise eligible encounters were removed for hospice care within 90 days of the eligible encounter, and 248 were removed for palliative care within 90 days of the eligible encounter (**Table 13**). As a non-interoperable system with limited information on external care services, hospice and palliative care information are combined in Site 2.

	Site 1 Before Exclusion	Site 1 After Exclusion	% lost	Site 2 Before Exclusion	Site 2 After Exclusion	% lost
Hospice Encounter	3572	3535	0.01%	N/A	N/A	N/A
Palliative Care	3572	3324	0.07%	842	632	24.94%

Table 13: Encounters Before and After Applying Measure Exclusions

[Response Ends]

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

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

[Response Begins]

Hospice and palliative events demonstrated mixed frequency across sites. Despite this variation, these exclusions are warranted given the different care and mobility goals of individuals facing long-term or terminal illnesses in preventing, identifying, and treating VTE events.

[Response Ends]

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

[Response Begins]

No risk adjustment or risk stratification

Other approach to address risk factors (specify)

[Other approach to address risk factors (specify) Please Explain]

Exclusions (hospice, palliative care)

[Response Ends]

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

[Response Begins]

N/A, no statistical risk models used.

[Response Ends]

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

[Response Begins]

This measure assesses the rate of encounters where a VTE is diagnosed >24 hours following a primary care visit and VTE symptoms have been documented in the EHR clinical notes. This means that at the time of the primary care visit, the clinician had the information necessary to diagnose a VTE via self-reported patient symptoms but did not make this diagnosis in a timely manner, which is dangerous for the treatment and management of VTEs.

In literature, there are minimal to no differences in hospital length of stay (LOS) between men and women hospitalized for VTE, and no significant differences in mortality between men and women diagnosed with VTE (Marshall et al., 2017; Mansour et al., 2017). Risk of VTE is associated with older age (Anderson et al., 1991; Silverstein et al., 1998; Gillum et al., 1987). African American race is associated with higher rates of VTE complications compared to white race (Aujesky et al., 2007). Although there are some disparities in the individuals who experience VTEs, there should not be social disparities in the delayed diagnosis of VTE following the onset of symptoms noted by a physician. For this measure, risk adjustment based on patient characteristics would establish

a lower standard of care for individuals with risk-adjusted characteristics as they are unrelated to delayed diagnosis. The goal of this measure is to quantify and reduce delayed VTE events, risk adjustment would mask the rate of delayed events among vulnerable populations and is not beneficial for this measure.

Additionally, VTEs are rare and dangerous events. Risk adjustment would impose sample size minimums at the clinician group level which would result in high numbers of group-level dropout and limit the monitoring potential of the measure. Stratification by patient risk factors would impose similar limitations. By not risk-adjusting the measure, we can use the model predictors to calculate expected rates for clinician groups who use the eCQM to compare against the observed rate. The observed over-expected ratio allows us to define clinician groups who are performing better than, worse than, or similar to expected rates based on their patient populations. In conversations with our Technical Expert Panel (TEP), we found that this measure would be more meaningful to patients and providers with the use of predictors than with the inclusion of a risk adjustment model. Currently, there is no national-level monitoring system to assess VTE events, in addition to benefits within a payment program, this measure could serve as the first passive monitoring system to assess the delayed diagnosis of VTE at the national level.

References:

1. Marshall, A.L., Bartley, A.C., Ashrani, A.A., Pruthi, R.K., Durani, U., Gonsalves, W.I., Kapoor, P., Hashmi, S.K., Siddiqui, M.A. and Go, R.S., 2017. Sex-based disparities in venous thromboembolism outcomes: A National Inpatient Sample (NIS)-based analysis. *Vascular Medicine*, 22(2), pp.121-127.
2. Mansour, S., Alotaibi, G., Wu, C., Alsaleh, K. and McMurtry, M.S., 2017. Sex disparities in hospitalization and mortality rates for venous thromboembolism. *Journal of Thrombosis and Thrombolysis*, 44(2), pp.197-202.
3. Anderson FA JrWheeler HBGGoldberg RJ et al. A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism: The Worcester DVT study. *Arch Intern Med* 1991;151933- 938
4. Silverstein MD Heit JA Mohr DN Petterson TM O'Fallon WM Melton LJ III Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population based study. *Arch Intern Med* 1998;158585- 593
5. Gillum RF Pulmonary embolism and thrombophlebitis in the United States, 1970-1985. *Am Heart J* 1987;1141262- 1264
6. Aujesky, D., Long, J.A., Fine, M.J. and Ibrahim, S.A., 2007. African American race was associated with an increased risk of complications following venous thromboembolism. *Journal of clinical epidemiology*, 60(4), pp.410-416.

[Response Ends]

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

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

Face validity established by a technical expert panel

[Response Ends]

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

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

[Response Begins]

N/A, no statistical risk models or stratification used

[Response Ends]

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

[Response Begins]

N/A, no statistical risk models or stratification used

[Response Ends]

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

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

[Response Begins]

N/A, no statistical risk models or stratification used

[Response Ends]

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

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

[Response Begins]

N/A, no statistical risk models or stratification used

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

N/A, no statistical risk models or stratification used

[Response Ends]

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

[Response Begins]

N/A, no statistical risk models or stratification used

[Response Ends]

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

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

[Response Begins]

N/A, no statistical risk models used

[Response Ends]

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

[Response Begins]

N/A, no statistical risk models or stratification used

[Response Ends]

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

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

[Response Begins]

N/A, no statistical risk models or stratification used

[Response Ends]

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

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

[Response Begins]

N/A, no statistical risk models used

[Response Ends]

3. Feasibility

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

3.05. Complete and attach the [NQF Feasibility Score Card](#).

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]

[Response Ends]

4. Usability and Use

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

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

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

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

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

[Response Begins]

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

5. Comparison to Related or Competing Measures

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

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

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

NOTE: If there are no related measures, please select N/A.

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

NOTE: If there are no competing measures, please select N/A.

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

Provide analyses when possible.

[Response Begins]

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.:

Available in attached file

Attachment: 3749e_DOVE_nqf_ecqm_feasibility_final_scorecard.xlsx

Attachment: 3749e_Value Sets.zip

Contact Information

Measure Steward (Intellectual Property Owner): Brigham and Women's Hospital

Measure Steward Point of Contact: Bowen, Mica, mrbowen@bwh.harvard.edu

Measure Developer if different from Measure Steward: Brigham and Women's Hospital

Measure Developer Point(s) of Contact: Dykes , Patricia, pdynes@bwh.harvard.edu

Bowen, Mica, mrbowen@bwh.harvard.edu

Additional Information

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

[Response Begins]

Available in attached file

[Response Ends]

Attachment: 3749e_DOVE_nqf_ecqm_feasibility_final_scorecard.xlsx

Attachment: 3749e_Value Sets.zip

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

Describe the members' role in measure development.

[Response Begins]

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]