Measure Worksheet (MEW-PA-New)

Partnership for Quality Measurement

Click here for Pre-Evaluation Public Comments

Click here for Measure Specifications

Content

Brief Measure Information

CBE #: 3749e

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 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: VTE is a serious, preventable public health problem affecting approximately 300,000–600,000 individuals in the U.S. each year and requires timely and adequate treatment (Beckman, et al. 2010). VTE consists of pulmonary embolism and deep vein thrombosis and its 30-day mortality rate is up to 23% (Tagalakis, et al. 2013; Nijekuter, et al. 2007). Because signs and symptoms of VTE are non-specific, timely recognition of VTE is difficult, and missed VTE diagnosis is common. Two classic studies of necropsies in large hospitals found that 9%-12% had VTE and 84%-91% were undiagnosed at the time of death (Karwinksi, et al. 1989, Carvalho Bricola, et al. 2013).

In addition to concerns over patient safety, VTE events are costly to healthcare systems. Ruppert et al. (2011) estimated that VTE complications ranged from $426-$41,133 across literature and represent a financial burden on healthcare systems. Preventing VTE events prevents resulting adverse events from occurring, meaning that this eCQM has the potential to save thousands of dollars in avoided healthcare costs at the patient level.

In 2008, the U.S. Surgeon General declared VTE a public health emergency and issued an official call for action to prevent DVT and PE. The surgeon general warned that while morbidity and mortality related to other deadly cardiovascular diseases have greatly improved over the past decade, VTE-related outcomes have not improved and without extensive efforts this problem will worsen as the population ages (DHHS, 2008).

In 2019, the American Society of Hematology published VTE diagnosis guidelines to provide an evidence-based strategy to efficiently

Content
evaluate patients (Anderson et al., 2019). The goal of these guidelines is to improve diagnostic accuracy by assisting providers with evaluating patients with suspected VTE while reducing unnecessary and more invasive testing (Lim, et al. 2018). While routine use of guidelines in primary care would likely reduce the number of missed or delayed VTE diagnoses, integration into practice is challenging as VTE symptoms are nonspecific and often present as symptoms consistent with an underlying chronic illness. Strategies such as measurement of diagnostic performance are needed to assist primary care providers with adopting VTE diagnosis guidelines and routinely using them in clinical practice. Currently there is no way to measure VTE diagnostic performance. Metrics are needed to quantify suboptimal VTE diagnostic performance, improve early recognition of VTE symptoms, and ultimately reduce unfavorable VTE outcomes. 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 eCQM. 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. This measure has the potential to lower health care costs associated with VTE by providing ongoing patient outcome data that can be used to improve VTE diagnostic performance and to reduce complications associated with delayed diagnosis and treatment. Citations: 1. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous Thromboembolism: A Public Health Concern. Am J Prev Med. 2010 Apr 1;38(4, Supplement):S495–501. 2. Tagalakis V, Patenaude V, Kahn SR, Suissa S. Incidence of and Mortality from Venous Thromboembolism in a Real-world Population: The Q-VTE Study Cohort. Am J Med. 2013 Sep 1;126(9):832.e13-832.e21. 3. Nijkeuter, M., Söhne, M., Tick, L.W., Kamphuisen, P.W., Kramer, M.H., Laterveer, L., van Houten, A.A., Kruip, M.J., Leebeek, F.W., Büller, H.R. and Huisman, M.V., 2007. The natural course of hemodynamically stable pulmonary embolism: clinical outcome and risk factors in a large prospective cohort study. Chest, 131(2), pp.517-523. 4. Karwinski B, Svendsen E. Comparison of clinical and postmortem diagnosis of pulmonary embolism. J Clin Pathol. 1989 Feb;42(2):135-9. doi: 10.1136/jcp.42.2.135. PMID: 2921354; PMCID: PMC1141815. 5. Carvalho Bricola, S.A.P., Paiva, E.F., Lichtenstein, A., Gianini, R.J., Duarte, J.G., Shinjo, S.K., Eluf-Neto, J. and Arruda Martins, M., 2013. Fatal pulmonary embolism in hospitalized patients: a large autopsy-based matched case-control study. Clinics, 68, pp.679-685. 6. Ruppert A, Steinle T, Lees M. Economic burden of venous thromboembolism: a systematic review. Journal of medical economics. 2011 Jan 1;14(1):65-74. 7. The Surgeon General's call to action to prevent deep vein thrombosis and pulmonary embolism. United States Department of Health and Human Services. Office of the Surgeon General (US) CTI - Publications and Reports of the Surgeon General; 2008. 8. Anderson, D.R., Morgano, G.P., Bennett, C., Dentali, F., Francis, C.W., Garcia, D.A., Kahn, S.R., Rahman, M., Rajasekhar, A., Rogers, F.B. and Smythe, M.A., 2019. American Society of Hematology 2019 guidelines for management of venous

Content
thromboembolism: prevention of venous thromboembolism in surgical hospitalized patients. Blood advances, 3(23), pp.3898-3944. 9. Lim, W., Le Gal, G., Bates, S.M., Righini, M., Haramati, L.B., Lang, E., Kline, J.A., Chasteen, S., Snyder, M., Patel, P. and Bhatt, M., 2018. American Society of Hematology 2018 guidelines for management of venous thromboembolism: diagnosis of venous thromboembolism. Blood advances, 2(22), pp.3226-3256
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 in the 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. **Table 1: VTE-related symptoms** Cough syncope shortness of breath foot pain foot numbness foot tingling foot redness foot swelling foot tenderness foot warmth hypotension tachycardia calf pain Calf numbness calf tingling calf redness

Content

calf swelling
calf tenderness
calf warmth
lightheadedness
hemoptysis
leg pain
leg numbness
leg tingling
leg redness
leg swelling
leg tenderness
leg warmth

sp.16. Denominator Exclusions: This electronica clinical quality measure (eCQM) excludes patients who received hospice or palliative care within 90 days of the eligible VTE event.

Measure Type: Intermediate Clinical Outcome

sp.28. Data Source: Electronic Health Records (EHR)

sp.07. Level of Analysis: Clinician: Group/Practice; Integrated Delivery System

IF Endorsement Maintenance—Original Endorsement Date: New measure

Most Recent Endorsement Date: New measure

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

IF this measure is paired/grouped, CBE#/title: N/A

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

Content
Staff Assessment: New Measure
Criterion 1: Importance to Measure and Report
1a. Evidence
1a. Evidence. The evidence requirements for a ***structure, process, or intermediate outcome*** measure are that it is based on a systematic review (SR) and grading of the body of empirical evidence in which the specific focus of the evidence matches what is being measured. For measures derived from a patient report, the evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.
The developer provides the following description for this measure: • This is a new intermediate clinical outcome measure at the clinician group practice and integrated service delivery levels that calculates the rate of delayed diagnosis of venous thromboembolism (VTE) in adults aged 18 years and older in the primary care setting. Delayed diagnosis is defined as 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 developer provides a logic model that connects a reduction in delay of diagnosis of VTE (DOVE) with outcomes such as reduction in adverse events associated with DOVE, reduction in healthcare costs, increased self-efficacy of physicians to address VTE symptoms. The principal mechanism the developer proposes for reduction of DOVE is a clinical decision support (CDS) tool that is designed to accompany this eCQM.
The developer provides the following evidence for this measure: • SR of the evidence specific to this measure? ☒ Yes ☐ No • Quality, Quantity, and Consistency of evidence provided? ☒ Yes ☐ No • Evidence graded? ☐ Yes ☒ No
Summary: • The developer summarizes systematic reviews (n=4) and guidelines (n=3) showing that VTE is consistently and substantially associated with deleterious health outcomes, including acute pulmonary embolism (APE), Chronic thromboembolic pulmonary hypertension (CTEPH), post-thrombotic syndrome, and death. • These sources and others the developers reference also demonstrate that failure or delay in diagnosis of VTE, through failure to provide timely treatment, can increase the risk of poor health outcomes, including death. Prompt treatment can reduce the

Content
risk of complications and mortality following from VTE. • Effective diagnostic methods exist, but symptoms of VTE can be non-specific and many cases are not diagnosed. In addition, there is the risk of diagnosing VTE inappropriately, leading to overtreatment and increased cost and risk.
Exception to evidence • N/A
Questions for the Standing Committee: • *What is the relationship between this measure and patient outcomes?* • *How strong is the evidence for this relationship?* • *Is the evidence directly applicable to the intermediate clinical outcome being measured?* • *Do primary care physicians usually have the capacity to confirm a VTE diagnosis during a visit? Is the CDS the developers propose to develop sufficient for this need?*
Guidance From the Evidence Algorithm Box 1: No → Box 3: No grading of evidence → Box 7: Yes → Box 8: Yes → Box 9: Moderate The highest possible rating is moderate.
Preliminary rating for evidence: ☒ Moderate ☐ Low ☐ Insufficient
1b. Gap in Care/Opportunity for Improvement and Disparities
1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement. • Performance data are reported for two sites: Boston, MA (01/06/2016 - 12/31/2021) and Lexington, KY (12/01/2016 - 12/31/2020). Boston performance scores are reported at the clinician group level in deciles and Lexington scores are reported for the facility overall. **Site 1: Boston, MA area** Sample size (total denominator): 3,591 eligible patient encounters Mean performance score: 72.6% Standard deviation: 21.52% Min performance score: 0% Max performance score: 100% Interquartile range for performance scores: 40% Performance scores by decile:

Content

Decile	DOVE Rate (%)
10	33.33
20	50
30	64.17
40	66.67
50	74.34
60	84.62
70	100
80	100
90	100
100	100

Data source: A total of 214 primary care clinician groups across the Site 1 healthcare system were included in the analysis. 5,514 patient encounters were assessed, 3,591 of which met the inclusion criteria for the measure.
**Clinician groups with ≥1 encounter were included in performance score data **

Site 2: Lexington, KY area
Sample size (total denominator): 245
Mean performance score: 77.14%
Standard deviation and other variance measures: N/A, rate calculated at the organization level (single entity)

Disparities

A sub-analysis was performed to assess disparities by social determinant of health variables in Site 1 patients (Boston, MA; Table 3). A T-test assuming equal variances was performed to assess if there was significant variation between subsamples. No significant differences were found across patients by race, ethnicity, sex, insurance, and age, meaning there were no significant differences in delayed VTE diagnosis rate by these patient characteristic.

Table 3: Sub Analysis of Site 1 Patients by Race, Ethnicity, Sex, Insurance, and Age

Demographics	**Denominator (column %)**	**Numerator (column %)**	**Dove Rate**	**P Value**
Total sample:	3591	2607	72.60%	
Race:				
Black:	312 (8.69)	213 (8.17)	68.27%	**0.841**

Content

Demographics	Denominator (column %)	Numerator (column %)	Dove Rate	P Value
White:	2945 (82.01)	2142 (82.16)	72.73%	
Other:*	334 (9.30)	252 (9.67)	75.45%	
Ethnicity:				
Hispanic:	233 (6.49)	172 (6.60)	73.82%	**0.548**
non-Hispanic:	3290 (91.62)	2389 (91.64)	72.61%	
Missing/declined:	68 (1.89)	46 (1.76)	67.65%	
Sex:				
Female:	1847 (51.43)	1346 (51.63)	72.87%	**0.979**
Male:	1744 (48.57)	1261 (48.37)	72.31%	
Insurance:				
Public:	1969 (54.83)	1472 (56.46)	74.76%	**0.387**
Private:	1611 (44.86)	1128 (43.27)	70.02%	
Other:**	11 (0.31)	7 (0.27)	63.64%	
Age:				
<65:	1509 (42.02)	1047 (40.16)	69.38%	**0.888**
≥65:	2082 (57.98)	1560 (59.84)	74.93%	

*Other racial category includes Asian, American Indian, Alaska Native, and race self-reported as “other”
**Other insurance category includes free care from the hospital and self-pay

Limited disparities evidence from the literature:

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 disparities in VTE prevalence by African American race are apparent across published literature, this eCQM specifically targets delayed diagnosis of VTE after symptoms are recorded in a primary care visit.

Questions for the Standing Committee:

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

Content
☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criteria 2: Scientific Acceptability of Measure Properties
Evaluators: Battelle Staff
2a. Reliability: Specifications and Testing
2a1. Specifications require the measure, as specified, to produce consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented. • The submitted measure specification follows established technical specifications for electronic clinical quality measures (eCQMs) (Quality Data Model [QDM], health quality measure format [HQMF], and Clinical Quality Language [CQL]) as indicated in subcriterion 2a1. • The submitted measure specification is fully represented and is not hindered by any limitations in the established technical specifications for eCQMs.
2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers. • The Site 1 sample included a total of 214 primary care sites. 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 Site 2 encounters did not meet inclusion criteria, compared to 34.87% in Site 1). • Accessing care across sites is a limitation of eCQMs in non-interoperable systems and is not limited to the DOVE eCQM. Based on testing in Site 2, the developer determined that the measure would be most meaningful when used within an integrated care delivery network. The developer is currently testing the measure in a third site that is an integrated care delivery network.
Specifications: • eCQM is specified using the latest industry-accepted eCQM technical specifications: HQMF, QDM, CQL, and value sets vetted through the National Library of Medicine's (NLM) Value Set Authority Center (VSAC).

Content
Reliability Testing: • Reliability testing conducted at the Patient/Encounter Level: o The developer conducted inter-abstractor NLP Algorithm Accuracy: used to extract VTE-related symptoms from the EHR clinical notes and inter-abstractor VTE Phenotyping Algorithm Accuracy: used to determine VTE diagnoses for denominator inclusion. o NLP Algorithm Accuracy: o 26 rounds of chart review were conducted with patients who had a VTE diagnosis (case cohort). 5 rounds of chart reviews were conducted with patients without VTEs (control cohort). Each round averaged 676 sentences of clinical notes. o Inter rater reliability of the NLP algorithm was 100% across 30 clinical encounter notes in two sites. The final Kappa was 1.00, representing almost perfect agreement. o VTE Phenotyping Algorithm Accuracy: o Positive Predictive Value (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 the 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, the algorithm's PPV was 95.80%. o Negative Predictive Value (NPV): Of the 500 randomly reviewed patients selected to determine the pipeline's NPV, the developer found that no patients had a true VTE. o Sensitivity and Specificity: Using the true positive, false positive, true negative, and false negative rates, the algorithm produced sensitivity and specificity rates of 100% and 95.69%, respectively. • Reliability testing conducted at the Accountable Entity Level: o A signal to noise analysis was conducted for the 15 largest practices at the individual clinician level (n=29) and the clinician group level (n=15) in Site 1. o A signal to noise analysis estimates the proportion of overall variability explained by the differences between measured entities (between individual clinicians, between clinician groups). o A minimum sample size of 10 encounters was required for the signal to noise analysis. o At the clinician group/practice-level, 15 groups from Site 1 were sampled and the median signal-to-noise (SNR) statistical result was 0.5393 (95% CI; 0.4166, 0.6619). The minimum SNR was 0.37512 and the maximum was 0.9489. o At the individual clinician level, 29 groups from Site 1 were sampled and the median signal-to-noise statistical result was 0.19054 (95% CI: 0.1544, 0.1983). The minimum SNR was 0.1225, the maximum was 0.3262. o There are no results for reliability testing at the integrated delivery system level.

Content
Questions for the Standing Committee regarding reliability: • *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*
Guidance From the Reliability Algorithm Box 1: Yes → Box 2: Yes → Box 4: No testing at integrated delivery system-level → Box 8: Yes → Box 9: Yes → Box 10: Moderate Highest possible rating is moderate.
Preliminary rating for reliability: ☒ Moderate ☐ Low ☐ Insufficient
2b. Validity: Validity Testing; Exclusions; Risk Adjustment; Meaningful Differences; Comparability; Missing Data
2b2. Validity testing should demonstrate that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b2-2b6. Potential threats to validity should be assessed/addressed.
 	
Validity Testing
• Validity testing conducted at the Patient/Encounter Level:
  ○ Following the manual chart review of 30 patients from Site 1, 22 patients were sorted into the denominator, 9 of the 22 patients from the denominator were included in the numerator, and 8 patients were excluded from the measure.
  ○ Manual chart review and the eCQM had 100% agreement (kappa = 1.0, PPV=100%, NPV=100%), demonstrating strong validity and agreement in the eCQM.
• Validity testing conducted at the Accountable Entity Level:
  ○ The developer conducted face validity and empirical validity testing.
  ○ Face validity:
  ○ The developer convened a technical expert panel (TEP) consisting of 6 members; three clinicians, one EHR expert, and two patient perspectives.
  ○ During a July 2022 meeting, the TEP was presented final measure specifications, initial rate calculations, and information on delayed diagnosis of VTE across literature. The TEP also had an opportunity to discuss questions and provide feedback to the measure development team.
  ○ A formal face validity vote was conducted via an online survey (Google Poll) that was sent to the TEP members by email after the presentation and discussion. Only TEP members who were present for this meeting were eligible to

Content

participate in the face validity vote.
    - The survey asked the following face validity question: “The VTE Diagnostic Delay in Primary Care eCQM, as specified, can be used to distinguish good form poor clinician group-level quality related to patient safety”. TEP members were blinded to each other’s responses, but were told the final face validity vote after all eligible members had voted.
    - The final vote was 5/5 (100%) in agreement with the voting statement among present members. 1 member was absent and did not vote.
  - Empirical Validity:
    - A random half split correlation was conducted at the clinician group level in Site 1, with 15 clinician groups included in the analysis. A minimum of 20 encounters for each eligible clinician was required.
    - 15 clinician groups were included in the analyses from Site 1. 1,168 encounters from 15 clinician groups were included in the test sample, 1,177 encounters from 15 clinician groups were included in the validation sample. The DOVE rate in the test sample was 72.52%, the DOVE rate in the validation sample was 75.02%. P values were calculated for encounter-level demographics, no variables were significantly different between test and validation groups.
    - The Spearman correlated was 0.7817 (95% CI: 0.4372, 0.9429). The ICC in the test sample was 0.0174 (95% CI: 0.0042, 0.6701). The ICC in the validation sample was 0.0262 (95% CI: 0.0086, 0.2654).

- The Feasibility Scorecard indicated that none of the measure data elements present issues with accuracy.
  - Data elements:
    - Encounter, Performed: Office Visit
    - Encounter, Performed: Outpatient
    - Encounter, Performed: Inpatient Encounter
    - Participant Roles
    - Diagnostic Study, Performed: Imaging Related to VTE
    - Diagnosis: VTE Diagnoses
    - Documented VTE Symptom
    - Medication, Order: Anticoagulant Medications
    - Intervention, Order: Hospice Care Ambulatory
    - Encounter, Performed: Palliative Care Encounter
    - Date of birth
    - Race
    - ONC Administrative Sex"Patient Characteristic Sex: ONC AdministrativeSex":

Content
Exclusions • The measure excludes patients who are in hospice or palliative care within 6 months of a VTE event. Impact to the measure is considered to be minimal.
Risk Adjustment • The measure is not risk-adjusted or stratified.
Meaningful Differences • To assess clinically and practically meaningful differences in performance measure scores among samples, the developer stratified clinician groups by encounter sample sizes into five cohorts and assessed overall DOVE rate and range. • 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. • Rates between cohorts ranged from 65.78%-77.14%, and variation within cohorts is seen in each cohort rate range. Although it is not clear whether the scores are statistically significant.
Missing Data • The measure is based on the presence of encounters, tests, and diagnoses and Rx orders, so missing data cannot be assessed and pose minimal threat to validity. • The developer states that all data elements required for encounter inclusion in the measure and measure calculation are commonly available within the EHR.
Comparability • The measure only uses one set of specifications for this measure.
Questions for the Standing Committee regarding validity: • *The committee should clarify with the developer if this measure is at the clinician-group practice level only, or if integrated delivery systems should also be considered, as no testing was provided for this level of analysis.* • *The committee should clarify with the developer if the split-half analysis is a true test of validity of the measure score.* • *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk adjustment approach, etc.)?* • *Are the accuracy issues that are captured in the Feasibility Scorecard substantial enough to impact the validity of these data elements?*
Guidance From the Validity Algorithm Box 1: Yes → Box 2: No → Box 3: Yes → Box 4: Moderate The highest possible rating is moderate, as the split-half testing of the measure score does to provide an indication whether the score is a true indicator of health care quality.

Content
Preliminary rating for validity: ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Criterion 3. Feasibility
3. Feasibility is the extent to which the specifications, including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement. • There are no fees associated with the use of this measure. The NLP algorithm that can be used if the VTE symptoms are not available as structured data, which includes an information overview on how the algorithm is used and instructions on implementation. • Using a simulated data set, the submission demonstrates that the evaluation of 100 percent of the measure logic can be automated.
Questions for the Standing Committee: • *Are the required data elements routinely generated and used during care delivery?* • *Are the required data elements available in electronic form (e.g., EHR or other electronic sources)?* • *Is the data collection strategy ready to be put into operational use?* • *For data elements assessed to have feasibility issues, does the developer present a credible, near-term path to electronic collection?*
Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**
• High – there are no feasibility issues identified in the scorecard and 100 percent coverage in simulated data unit tests (BONNIE) • Moderate – all identified feasibility issues have a core plan to address the issues and 100 percent coverage in simulated data unit tests (BONNIE) • Low – identified feasibility issues do not have a plan to address the issues and 100 percent coverage in simulated data unit tests (BONNIE) • Insufficient – no feasibility scorecard or simulated data unit tests (BONNIE) included in the submission
Criterion 4: Use and Usability
4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

Content
4a. Use evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.
4a1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.
Current uses of the measure Publicly reported? ☐ Yes ☒ No Current use in an accountability program? ☐ Yes ☒ No ☐ UNCLEAR Planned use in an accountability program? ☒ Yes ☐ No ☐ N/A
Accountability program details • This new eCQM is currently not in use. However the developer states that the measure will be submitted in May 2023 for potential inclusion as a Merit-based Incentive Payment System (MIPS) measure for the CMS Quality Payment Program (QPP).
4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: (1) Those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; (2) Those being measured, and other users have been given an opportunity to provide feedback on the measure performance or implementation; and (3) This feedback has been considered when changes are incorporated into the measure.
Feedback on the measure provided by those being measured or others • The developer performed several forms of stakeholder engagement throughout the measure development and testing process, including feedback from a technical expert panel (TEP) (described in validity testing), focus groups with providers and patients who had survived a VTE event, and public comment solicitation. • The developer used the feedback from our TEP to refine specifications like defining VTE by the presence of ICD-10, RxNorm, and CPT codes, and adding additional symptoms of VTE in the primary care setting for the NLP algorithm. Additional VTE symptoms were added to its symptom list based on feedback from focus groups with providers. The TEP and provider interviews were used to validate the measure specifications from a clinician and EHR vendor perspective.
Questions for the Standing Committee: • *Can the performance results be used to further the goal of high quality, efficient healthcare?*

Content
Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**
4b. Usability (4b1. Improvement; 4b2. Benefits of measure)
4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.
4b1 Improvement. Progress toward achieving the goal of high quality, efficient healthcare for individuals or populations is demonstrated. **Improvement results** • This is a new measure, therefore, the developer states that there is no information available on performance improvement. This measure is not currently used in a program, but the developer emphasizes that the primary goal of the measure is to provide information necessary for public reporting and quality improvement.
4b2. Benefits versus harms. The benefits of the performance measure in facilitating progress toward achieving high quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists). **Unexpected findings (positive or negative) during implementation** • The devleoper states that there were no identified unexpected findings during testing and the measure is not currently in use. **Potential harms** • No harms identified, as not currently in use.
Additional Feedback: • None
Questions for the Standing Committee: • *How can the performance results be used to further the goal of high quality, efficient healthcare?* • *Do the benefits of the measure outweigh any potential unintended consequences?*
Preliminary rating for Usability and Use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Content
Criterion 5: Related and Competing Measures
Related Measures • The developer states that there are no related or competing measures.
Harmonization • N/A

QUALITY MEASURE SUBMISSION FORM

Version: 1.0; Generated: 13 April 2023

Introduction

Thank you for your interest in submitting a measure to Battelle for possible endorsement.

What criteria are used to evaluate measures? Measures are evaluated on standardized criteria: importance to measure and report, scientific acceptability of measure properties, feasibility, usability and use, and related and competing measures. For your measure to be evaluated against these measure evaluation criteria, you must complete the measure submission form.

Why do I have to complete a form? Due to the volume and/or complexity of proposed measures, Battelle provides measure information to committee reviewers in a standardized format to facilitate their evaluation of whether the measure meets the measure evaluation criteria. This form allows the measure steward to present information demonstrating that the proposed measure meets endorsement criteria.

What is on the form? The information requested in this form is directly related to the measure evaluation criteria.

Can't I just submit our files for consideration? No. Measures must be submitted through the online form to be considered for the Spring 2023 cycle. Requested information should be entered directly into this form and as well as any necessary or required attachments.

Can I submit additional details and materials? Additional materials will be considered only as supplemental. Do NOT rely on material provided in an appendix to provide measure specifications or to demonstrate meeting the criteria. The core information needed to evaluate the measure should be provided in the appropriate submission form fields and required attachments. Please contact PQMsupport@battelle.org regarding questions about submitting supplemental materials.

What do I do first? If you have started a new submission by answering five qualifying questions, you may proceed to the “Previous Submission Information" tab to continue with your submission. The “Conditions” tab will list the conditions that must be met before your proposed measures may be considered and evaluated for suitability as endorsed voluntary consensus standards. You are asked to acknowledge reading and accepting the conditions.

Can I make changes to a form once I have submitted it? No. Once you submit your measure, you will NOT be able to return to this submission form to make further revisions. You will need to contact project staff.

What if I need additional help? Please contact the project staff at PQMsupport@battelle.org if you have questions regarding the information requested or submitting supplemental materials.

NOTE: All measure submissions should be 508-compliant. Refer to the Checklist for Developer 508 Guidelines (PDF) to ensure all guidelines apply to all parts of your submission, including all fields and attachments used within the measure submission form.

Please email us at PQMsupport@battelle.org if you experience technical difficulties using the online submission form.

Thank you for your interest in submitting measures to Battelle.

Previous Submission Information (1 – 4)

1) Select whether this measure was previously submitted to the prior consensus-based entity (the National Quality Forum [NQF]) and given an identifying number.

☐ Previously submitted to NQF
☒ New measure, never submitted.

2) Provide the measure number of the previously submitted measure.
N/A, not previously submitted.

3) If the measure has an electronic clinical quality measure (eCQM) version, provide the measure number of the previously submitted measure.
N/A, not previously submitted.

4) If this eCQM has a registry version, provide the measure numbers of the previously submitted measure.
N/A, not previously submitted.

Conditions (1 - 2)

Several conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards. If any of the conditions are not met, the measure will not be accepted for consideration.

A. A Measure Steward Agreement is signed or the steward is a government organization. (All non-government organizations must sign a Measure Steward Agreement.) For more information about completing a Measure Steward Agreement, please go to: Endorsement | Partnership for Quality Measurement (p4qm.org) and follow the instructions.

B. The measure owner/steward verifies there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every three years.

C. The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient healthcare.

D. The measure is fully specified and tested for reliability and validity.

E. The measure developer/steward attests that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate.

F. The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided.

1) Check if either of the following apply.

☒ Proprietary measure or components (e.g., risk model, codes)
☐ Proprietary measure or components with fees
☐ None of the above

2) Check the box below to agree to the conditions listed above.

☒ I have read and accept the conditions as specified above

Specifications: Maintenance Update (spma.01 - spma.02)

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

☐ No
☐ Yes
N/A, new measure

spma.02) Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.
No changes, new measure.

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

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

Measure Specifications (sp.01 - sp.32)

sp.01) Provide the measure title.
Measure titles should be concise yet convey who and what is being measured.

Diagnostic Delay of Venous Thromboembolism (DOVE) in Primary Care

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).

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 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.

sp.03) Provide a rationale for why this measure must be reported with other measures to appropriately interpret results.

N/A, measure is not paired nor a composite.

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

☐ Behavioral Health
☐ Behavioral Health: Alcohol, Substance Use/Abuse
☐ Behavioral Health: Anxiety
☐ Behavioral Health: Attention Deficit Hyperactivity Disorder (ADHD)
☐ Behavioral Health: Bipolar Disorder
☐ Behavioral Health: Depression
☐ Behavioral Health: Domestic Violence
☐ Behavioral Health: Other Serious Mental Illness
☐ Behavioral Health: Post-Traumatic Stress Disorder (PTSD)
☐ Behavioral Health: Schizophrenia
☐ Behavioral Health: Suicide
☐ Cancer
☐ Cancer: Bladder
☐ Cancer: Breast
☐ Cancer: Colorectal
☐ Cancer: Gynecologic

- ☐ Cancer: Hematologic
- ☐ Cancer: Liver
- ☐ Cancer: Lung, Esophageal
- ☐ Cancer: Prostate
- ☐ Cancer: Renal
- ☐ Cancer: Skin
- ☐ Cancer: Thyroid
- ☐ Cardiovascular
- ☐ Cardiovascular: Arrythmia
- ☐ Cardiovascular: Congestive Heart Failure
- ☐ Cardiovascular: Coronary Artery Disease
- ☐ Cardiovascular: Coronary Artery Disease (AMI)
- ☐ Cardiovascular: Coronary Artery Disease (PCI)
- ☐ Cardiovascular: Hyperlipidemia
- ☐ Cardiovascular: Hypertension
- ☐ Cardiovascular: Secondary Prevention
- ☐ Critical Care
- ☐ Critical Care: Assisted Ventilation
- ☐ Critical Care: Intensive Monitoring
- ☐ Dental
- ☐ Dental: Caries
- ☐ Dental: Tooth Loss
- ☐ Ears, Nose, Throat (ENT)
- ☐ Ears, Nose, Throat (ENT): Ear Infection
- ☐ Ears, Nose, Throat (ENT): Hearing
- ☐ Ears, Nose, Throat (ENT): Pharyngitis
- ☐ Ears, Nose, Throat (ENT): Tonsilitis
- ☐ Endocrine
- ☐ Endocrine: Calcium and Metabolic Bone Disorders
- ☐ Endocrine: Diabetes
- ☐ Endocrine: Female and Male Endocrine Disorders
- ☐ Endocrine: Hypothalamic-Pituitary Disorders
- ☐ Endocrine: Thyroid Disorders
- ☐ Eye Care
- ☐ Eye Care: Age-related macular degeneration (AMD)
- ☐ Eye Care: Cataracts
- ☐ Eye Care: Diabetic retinopathy
- ☐ Eye Care: Glaucoma
- ☐ Gastrointestinal (GI)
- ☐ Gastrointestinal (GI): Constipation

- ☐ Gastrointestinal (GI): Gall Bladder Disease
- ☐ Gastrointestinal (GI): Gastroenteritis
- ☐ Gastrointestinal (GI): Gastro-Esophageal Reflux Disease (GERD)
- ☐ Gastrointestinal (GI): Hemorrhoids
- ☐ Gastrointestinal (GI): Hernia
- ☐ Gastrointestinal (GI): Inflammatory Bowel Disease
- ☐ Gastrointestinal (GI): Irritable Bowel Syndrome
- ☐ Gastrointestinal (GI): Peptic Ulcer
- ☐ Genitourinary (GU)
- ☐ Genitourinary (GU): Benign Prostatic Hyperplasia
- ☐ Genitourinary (GU): Erectile Dysfunction/Premature Ejaculation
- ☐ Genitourinary (GU): Incontinence/pelvic floor disorders
- ☐ Genitourinary (GU): Prostatitis
- ☐ Genitourinary (GU): Urinary Tract Injection (UTI)
- ☐ Gynecology (GYN)
- ☐ Gynecology (GYN): Abnormal bleeding
- ☐ Gynecology (GYN): Endometriosis
- ☐ Gynecology (GYN): Infections
- ☐ Gynecology (GYN): Menopause
- ☐ Gynecology (GYN): Pelvic Pain
- ☐ Gynecology (GYN): Uterine fibroids
- ☐ Infectious Diseases (ID)
- ☐ Infectious Diseases (ID): HIV/AIDS
- ☐ Infectious Diseases (ID): Influenza
- ☐ Infectious Diseases (ID): Lyme Disease
- ☐ Infectious Diseases (ID): Meningococcal Disease
- ☐ Infectious Diseases (ID): Pneumonia and respiratory infections
- ☐ Infectious Diseases (ID): Sepsis
- ☐ Infectious Diseases (ID): Sexually Transmitted
- ☐ Infectious Diseases (ID): Tuberculosis
- ☐ Liver
- ☐ Liver: Viral Hepatitis
- ☐ Musculoskeletal
- ☐ Musculoskeletal: Falls and Traumatic Injury
- ☐ Musculoskeletal: Gout
- ☐ Musculoskeletal: Joint Surgery
- ☐ Musculoskeletal: Low Back Pain
- ☐ Musculoskeletal: Osteoarthritis
- ☐ Musculoskeletal: Osteoporosis
- ☐ Musculoskeletal: Rheumatoid Arthritis

☐ Neurology
☐ Neurology: Alzheimer's Disease
☐ Neurology: Autism
☐ Neurology: Brain Injury
☐ Neurology: Epilepsy
☐ Neurology: Migraine
☐ Neurology: Parkinson's Disease
☐ Neurology: Spinal Cord Injury
☐ Neurology: Stroke/Transient Ischemic Attack (TIA)
☒ Other (please specify here: **venous thromboembolism, primary care**)
☐ Palliative Care and End-of-Life Care
☐ Palliative Care and End-of-Life Care: Advanced Directives
☐ Palliative Care and End-of-Life Care: Amyotrophic Lateral Sclerosis (ALS)
☐ Palliative Care and End-of-Life Care: Hospice Management
☐ Palliative Care and End-of-Life Care: Inappropriate use of acute care services
☐ Palliative Care and End-of-Life Care: Pain Management
☐ Perinatal Health
☐ Perinatal Health: Labor and Delivery
☐ Perinatal Health: Newborn Care
☐ Perinatal Health: Post-Partum Care
☐ Perinatal Health: Preconception Care
☐ Perinatal Health: Prenatal Care
☐ Renal
☐ Renal: Acute Kidney Injury
☐ Renal: Chronic Kidney Disease (CKD)
☐ Renal: End Stage Renal Disease (ESRD)
☐ Renal: Infections
☐ Reproductive Health
☐ Reproductive Health: Family planning and contraception
☐ Reproductive Health: Infertility
☐ Reproductive Health: Male reproductive health
☐ Respiratory
☐ Respiratory: Acute Bronchitis
☐ Respiratory: Allergy
☐ Respiratory: Asthma
☐ Respiratory: Chronic Obstructive Pulmonary Disease (COPD)
☐ Respiratory: Dyspnea
☐ Respiratory: Pneumonia
☐ Respiratory: Sleep Apnea
☐ Surgery

☐ Surgery: Cardiac Surgery
☐ Surgery: Colorectal
☐ Surgery: Neurosurgery / Spinal
☐ Surgery: Orthopedic
☐ Surgery: Orthopedic Hip/Pelvic Fractures
☐ Surgery: Pediatric
☐ Surgery: Perioperative and Anesthesia
☐ Surgery: Plastic
☐ Surgery: Thoracic Surgery
☐ Surgery: Trauma
☐ Surgery: Vascular Surgery

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

☐ Access to Care
☐ Care Coordination
☐ Care Coordination: Readmissions
☐ Care Coordination: Transitions of Care
☐ Disparities Sensitive
☐ Health and Functional Status
☐ Health and Functional Status: Change
☐ Health and Functional Status: Nutrition
☐ Health and Functional Status: Obesity
☐ Health and Functional Status: Physical Activity
☐ Health and Functional Status: Quality of Life
☐ Health and Functional Status: Total Health
☐ Immunization
☒ Other (please specify here: **delayed diagnosis**)
☐ Person-and Family-Centered Care: Person-and Family-Centered Care
☐ Person-and Family-Centered Care: Workforce
☐ Primary Prevention
☐ Primary Prevention: Nutrition
☐ Primary Prevention: Tobacco Use
☒ Safety
☐ Safety: Complications
☐ Safety: Healthcare Associated Infections
☐ Safety: Medication
☐ Safety: Overuse
☐ Screening

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.

☒ Adults (Age >= 18)
☐ Children (Age < 18)
☐ Elderly (Age >= 65)
☐ Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid
☐ Populations at Risk: Individuals with multiple chronic conditions
☐ Populations at Risk: Veterans
☐ Women

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.

☐ Accountable Care Organization
☒ Clinician: Group/Practice
☐ Clinician: Individual
☐ Facility
☐ Health Plan
☒ Integrated Delivery System
☐ Other (please specify here:)
☐ Population: Community, County or City
☐ Population: Regional and State

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

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
☐ Ambulatory Care
☐ Behavioral Health
☐ Home Care
☐ Inpatient/Hospital
☐ Other (please specify here:)
☒ Outpatient Services
☐ Post-Acute Care

sp.09) Provide a Uniform Resource Locator (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".

None available

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

Attach the zipped output from the measure authoring tool (MAT) for eCQMs - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications). HQMF specifications are attached.

☐ HQMF specifications are NOT attached (Please explain).
HQMF specifications are attached.

sp.11) Attach the simulated testing attachment.

All eCQMs 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.
☒ Testing is attached
☐ Testing is NOT attached (please explain)

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 at PQMsupport@battelle.org. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

☒ Available in attached Excel or csv file
☐ No data dictionary/code table – all information provided in the submission form

For the question below: state the outcome/process being measured. Calculations of the risk-adjusted outcome measures 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.

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

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.

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**).

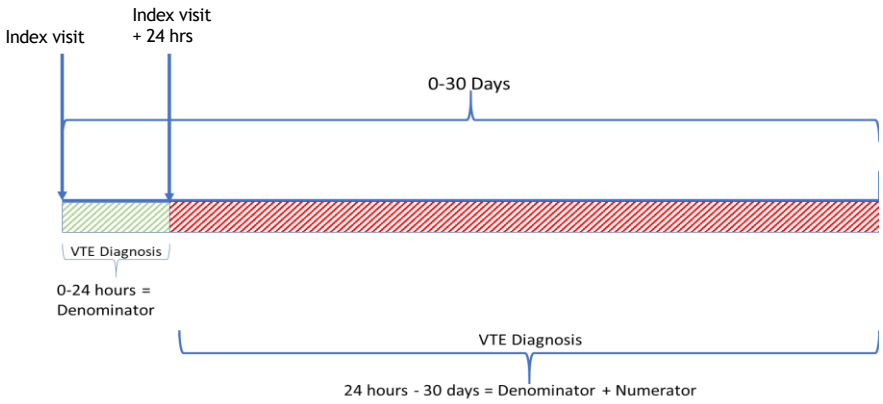

Figure 1: DOVE Numerator Inclusion Timeline

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

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.

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 in the 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.

Table 1: VTE-related symptoms

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

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

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.

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. A VTE diagnosis is defined using all three 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.
    - An encounter must have all three code types to indicate an eligible VTE event, encounters missing one of these code types will not be defined as a qualifying VTE.
4. Have no eligible VTE events within 6 months of the qualifying VTE event

sp.17) Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

This eCQM excludes patients who received hospice or palliative care within 90 days of the eligible VTE event.

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.

Patients who are on hospice or palliative care within 90 days of the eligible VTE encounter are excluded from this measure. The rationale for this exclusion is that these patients have different care goals than non-hospice or palliative care which may affect their VTE diagnosis. The value set for hospice and palliative care exclusions can be found in **Table 2**.

Table 2: Value Sets for Measure Exclusion Criteria

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

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.

N/A, measure is not stratified

sp.20) Is this measure adjusted for socioeconomic status (SES)?
☐ Yes
☒ No

sp.21) Select the risk adjustment type.
Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.
☒ No risk adjustment or risk stratification
☐ Statistical risk model
☐ Stratification by risk category/subgroup (specify number of risk factors)
☐ Other approach to address risk factors (please specify here:)

sp.22) Select the most relevant type of score.
Attachment: If available, please provide a sample report.
☐ Categorical, e.g., yes/no
☐ Continuous variable, e.g. average
☐ Count
☐ Frequency Distribution
☐ Non-weighted score/composite/scale
☐ Other (please specify here:)
☒ Rate/proportion
☐ Ratio
☐ Weighted score/composite scale

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.
☐ Better quality = Higher score
☒ Better quality = Lower score
☐ Better quality = Score within a defined interval
☐ Passing score defines better quality

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.

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. Include only the first eligible VTE encounter within a 6-month period per patient.

Step 2: define the denominator:
Identify qualifying VTE events. A qualifying VTE event is defined using the following three criterion within 30 days of the primary care visit (all must be present for measure inclusion):

- ICD-10 CM code for VTE.
- CPT codes of imaging 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.

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 multiple by 100. The measure is reported as a percentage: XX out of 100.

sp.25) Attach a copy of the instrument (e.g. survey, tool, questionnaire, scale) used as a data source for your measure, if available.

☐ Copy of instrument is attached.
☐ Copy of instrument is NOT attached (please explain).
N/A, no instrument.

sp.26) Indicate the responder for your instrument.

☐ Patient
☐ Family or other caregiver
☐ Clinician
☐ Other (specify)
N/A, no instrument.

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 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.

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

sp.28) Identify whether and how proxy responses are allowed.

sp.29) Survey/Patient-reported data.

Provide instructions for data collection and guidance on minimum response rate. Specify calculation of response rates to be reported with performance measure results.

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

☐ Assessment Data
☐ Claims
☐ Electronic Health Data
☒ Electronic Health Records
☐ Instrument-Based Data
☐ Management Data
☐ Other (please specify here:)
☐ Paper Medical Records
☐ Registry Data

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.

At the time of this submission, the DOVE 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. Currently we are conducing testing at a third large healthcare system in Pennsylvania.
Site 1 uses EPIC EHR, data were collected from 2016-2021. Data collected 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 2021.
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 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 have 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.

sp.32) Provide the data collection instrument.

☐ Available at measure-specific web page URL identified in sp.09
☐ Available in attached appendix in Question 1 of the Additional Section
☒ No data collection instrument provided

Importance to Measure and Report: Maintenance of Endorsement (1ma.01)

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

☐ Yes
☐ No
N/A, new measure.

Importance to Measure and Report: Evidence (Complete for Outcome Measures) (1a.01 - 1a.03)

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.

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.

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

Importance to Measure and Report: Evidence (Complete for Process Measures) (1a.03 - 1a.16)

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.

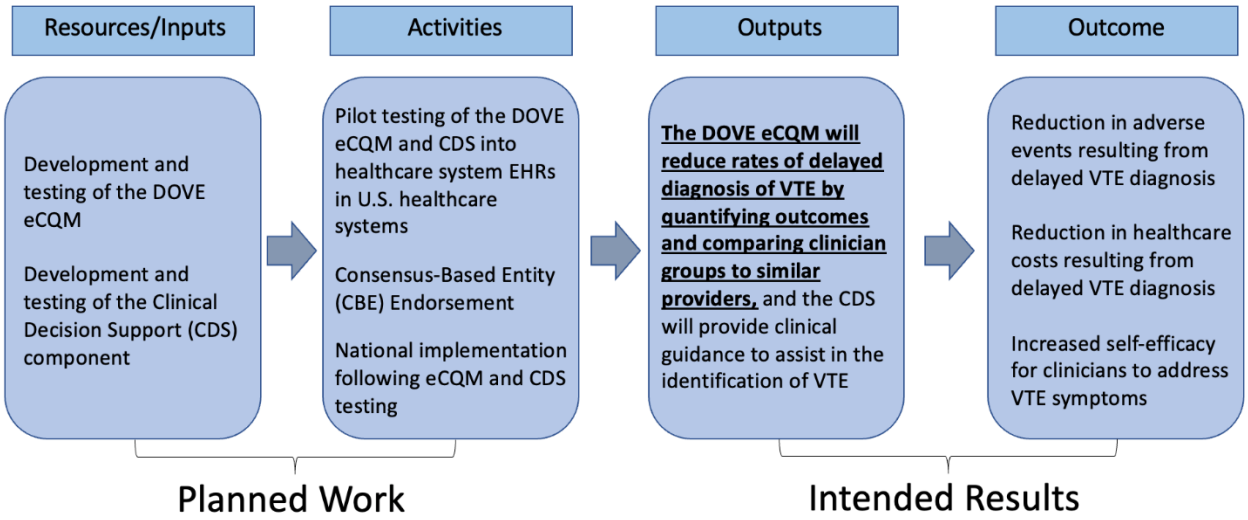

Figure 2: DOVE eCQM Logic Model

1a.02) Select the type of source for the systematic review of the body of evidence that supports the performance measure.
A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data.

☒ Clinical Practice Guideline recommendation (with evidence review)
☐ US Preventive Services Task Force Recommendation

☒ Other systematic review and grading of the body of evidence (e.g., Cochrane Collaboration, AHRQ Evidence Practice Center)
☐ Other (please specify here:)

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, you may add additional tables to the relevant sections. Please follow the 508 Checklist for tables.

Evidence - Systematic Reviews Table (Repeatable)
1a.03) Provide the title, author, date, citation (including page number) and URL for the systematic review.

Systematic Reviews:

1. Xiang, Qian, et al. "The predictive value of circulating microRNAs for venous thromboembolism diagnosis: A systematic review and diagnostic meta-analysis." *Thrombosis Research* 181 (2019): 127-134. https://www.sciencedirect.com/science/article/pii/S0049384819303172?casa_token=zUghOXCSG_cAAAAA:r811xc63C_3OxK8ZXBhcSGJaEvz1CTLdSFOYcUH2FOewUd0wtOQtLW59o8ddI5uWPxojGd8rNsM

2. Bhatt M, Braun C, Patel P, Patel P, Begum H, Wiercioch W, Varghese J, Wooldridge D, Alturkmani HJ, Thomas M, Baig M, Bahaj W, Khatib R, Kehar R, Ponnapureddy R, Sethi A, Mustafa A, Nieuwlaat R, Lim W, Bates SM, Lang E, Le Gal G, Righini M, Husainat NM, Kalot MA, Al Jabiri YN, Schünemann HJ, Mustafa RA. Diagnosis of deep vein thrombosis of the lower extremity: a systematic review and meta-analysis of test accuracy. Blood Adv. 2020 Apr 14;4(7):1250-1264. doi: 10.1182/bloodadvances.2019000960. PMID: 32227213; PMCID: PMC7160276**.** https://ashpublications.org/bloodadvances/article/4/7/1250/454151/Diagnosis-of-deep-vein-thrombosis-of-the-lower

3. Khan F, Tritschler T, Kimpton M, Wells PS, Kearon C, Weitz JI, Büller HR, Raskob GE, Ageno W, Couturaud F, Prandoni P, Palareti G, Legnani C, Kyrle PA, Eichinger S, Eischer L, Becattini C, Agnelli G, Vedovati MC, Geersing GJ, Takada T, Cosmi B, Aujesky D, Marconi L, Palla A, Siragusa S, Bradbury CA, Parpia S, Mallick R, Lensing AWA, Gebel M, Grosso MA, Thavorn K, Hutton B, Le Gal G, Fergusson DA, Rodger MA; MAJESTIC Collaborators. Long-Term Risk for Major Bleeding During Extended Oral Anticoagulant Therapy for First Unprovoked Venous Thromboembolism : A Systematic Review and Meta-analysis. Ann Intern Med. 2021 Oct;174(10):1420-1429. doi: 10.7326/M21-1094. Epub 2021 Sep 14. PMID: 34516270. https://www.acpjournals.org/doi/full/10.7326/M21-1094?casa_token=r92M7xlwErgAAAAA%3AOl_nWQcw-A7bugVpfz9UkxXE91skosW2ftkobBUIXSaX89U_4xZdD1h1mI-OO0nyvQrarxMqjK9VNXg

4. Becattini, Cecilia, et al. "Risk stratification of patients with acute symptomatic pulmonary embolism based on presence or absence of lower extremity DVT: systematic review and meta-analysis." *Chest* 149.1 (2016): 192-200. https://pubmed.ncbi.nlm.nih.gov/26204122/

Clinical guidelines:

5. New guidelines from the Thrombosis and Haemostasis Society of Australia and New Zealand for the diagnosis and management of venous thromboembolism https://www.thanz.org.au/documents/item/414

6. Qaseem A, Snow V, Barry P, Hornbake ER, Rodnick JE, Tobolic T, Ireland B, Segal J, Bass E, Weiss KB, Green L, Owens DK; Joint American Academy of Family Physicians/American College of Physicians Panel on Deep Venous Thrombosis/Pulmonary Embolism. Current diagnosis of venous thromboembolism in primary care: a clinical practice guideline from the American Academy of Family Physicians and the American College of Physicians. Ann Fam Med. 2007 Jan-Feb;5(1):57-62. doi: 10.1370/afm.667. PMID: 17261865; PMCID: PMC1783928. https://pubmed.ncbi.nlm.nih.gov/17261865/

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

Systematic Reviews:

1a. (Xiang et al., 2019) DVT occurs in deep veins generally located in the lower extremities, and the thrombus is only loosely connected to the vascular wall and may easily detach. Once the thrombus releases from the venous wall, it will enter the blood stream and cause embolism of lungs. When a thrombus attaches to the lungs, it can cause APE, and the repeated abscission of small clot fragments can result in CTEPH. To avoid serious or even fatal outcomes resulting from thromboembolisms such as APE, early diagnosis and preventative treatment should be a priority for high risk VTE patients. p. 128

- **Grade:** N/A, evidence review of a systematic review.

2a. (Bhatt et al., 2020) Deep vein thrombosis (DVT) of the lower extremities can be associated with significant morbidity and may progress to pulmonary embolism and postthrombotic syndrome. Early diagnosis and treatment are important to minimize the risk of these complications. P.1250
Early diagnosis and clinical intervention are important for managing DVT and minimizing adverse consequences, as well as to exclude the diagnosis in those who do not have the disease, thereby avoiding added costs and risks of anticoagulant therapy, p. 1251.

- **Grade:** N/A, evidence review of a systematic review.

3a. (Khan et al., 2021) Our comprehensive systematic review and meta-analysis demonstrates the considerable long-term risk and clinical impact of recurrent VTE while receiving extended anticoagulant therapy, with a
5-year cumulative incidence of 7% and a case-fatality rate of 5%. This information has important implications for clinical practice. Firstly, the diagnosis and management of recurrent VTE in patients receiving anticoagulation therapy remain difficult. P.2810

- **Grade:** evidence review of a systematic review.

4a. (Becattini et al., 2016) In this meta-analysis of 10 cohorts that included 7,868 patients with acute symptomatic PE, patients with concomitant ultrasound-detectable DVT had a 1.9-fold increased risk of short-term death compared with patients without DVT. Regardless of the length of follow-up, hemodynamic status, and study design, the studies showed relatively consistent results, without evidence of statistical heterogeneity. P. 198

- **Grade:** N/A, evidence review of a systematic review.

Clinical Guidelines:
5a. (THANZ, 2019) VTE can be fatal if untreated; long term morbidity includes post-thrombotic syndrome (PTS) and pulmonary hypertension. Symptoms of VTE are non-specific, and the diagnosis should actively be sought once considered. A diagnosis of VTE has an impact on subsequent pregnancies, oestrogen use, surgery, life insurance and, occasionally, long-haul travel.p.1.

- **Grade:** N/A, evidence review

5b. (THANZ, 2019) Clinical presentations of VTE are non-specific, and only about 20% of patients with clinically suspected VTE have it objectively confirmed. A misdiagnosis of VTE has significant implications including needless cessation of effective hormonal contraception in women and unnecessary ante- and post-partum injections of low molecular weight heparin, and in older patients, anticoagulation is associated with higher rates of major and fatal bleeding. Conversely, failure to diagnose VTE can result in fatal PE.p.2-3

- **Grade:** N/A, evidence review

6a. (Qaseem et al.,2007) Twenty-six percent of undiagnosed and untreated patients with pulmonary embolism will have a subsequent fatal embolic event, whereas another 26% will have a nonfatal recurrent embolic event that can eventually be fatal. Thus, the importance of early diagnosis to prevent mortality and morbidity associated with VTE cannot be overemphasized.

- **Grade:** N/A, evidence review

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

N/A, evidence sourced from evidence review sections, no grading.

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

Xiang et al., 2019: Quality of evidence was assessed by the diagnostic accuracy studies-2 (QUADAS-2) tool.
Bhatt et al., 2020: certainty of the evidence was assessed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework and ranged between low, moderate, and high.
Khan et al., 2021: N/A, evidence was not graded
Becattini et al., 2016: the Quality in Prognosis Studies tool was used to assess the quality of eligible studies across the following domains: study participation, follow up description, description of diagnosis of DVT, outcome defined and described appropriately, control of confounding, and analysis described appropriately. Rating ranged from low to high bias.
THANZ, 2019: The GRADE framework was used to determine the strength of the evidence, ranged from low to high.
Qaseem et al., 2007: N/A, evidence was not graded

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

N/A

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

Xiang et al., 2019: N/A, systematic review, no recommendation grading.
Bhatt et al., 2020: N/A, systematic review, no recommendation grading.
Khan et al., 2021: N/A, systematic review, no recommendation grading.
Becattini et al., 2016: N/A, systematic review, no recommendation grading.
THANZ, 2019: the GRADE system was used to determine the strength of recommendations, ranging from poor to strong.
Qaseem et al., 2007: N/A, recommendations were not graded.

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

Xiang et al., 2019: 12 studies were included.
Bhatt et al., 2020: 43 studies were included. Cohort and cross-sectional studies were included.
Khan et al., 2021: 27 studies were included, 14 randomized controlled trials (RCTs) and 13 cohort studies.
Becattini et al., 2016: 9 cohort studies were included.
THANZ, 2019: 51 studies were referenced in the guideline, including RCTs, observational studies, literature reviews, and systematic reviews.
Qaseem et al., 2007: 64 studies (RCT and observational) and 29 systematic reviews were included

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

Xiang et al., 2019: miR-132, among other novel diagnostic biomarkers for VTE, has shown to be both extensive and reliable.
Bhatt et al., 2020: serial ultrasound (US), proximal compression US, and whole-leg US to diagnose VTE had high sensitivity and specificity. D-dimer tests had high sensitivity but poor specificity.
Khan et al., 2021: Cumulative incidence of bleeding resulting from extended anticoagulation was consistently higher at 5 years post VTE than 2 years for major bleeding, intracranial bleeding, and fatal bleeding in all studies.
Becattini et al., 2016: Across studies, concomitant DVT was significantly associated with increased risk of death within 30 days of PE diagnoses for patients with acute symptomatic PE.
THANZ, 2019: N/A, guidelines did not include a systematic review and meta-analysis.
Qaseem et al., 2007: N/A, guidelines did not include a systematic review and meta-analysis.

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

Xiang et al., 2019: N/A, no harms identified
Bhatt et al., 2020: N/A, no harms identified
Khan et al., 2021: N/A, no harms identified
Becattini et al., 2016: N/A, no harms identified
THANZ, 2019: N/A, no harms identified
Qaseem et al., 2007: N/A, no harms identified

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

N/A, relevant systematic reviews have been included. No systematic reviews or guidelines specifically on the importance of proximal VTE diagnosis have been developed as it is well documented that delayed diagnoses are harmful to patients (see section 1a.14). As the harms of delayed diagnosis are clear, systematic reviews and guidelines are more focused on providing clinicians with appropriate diagnostic tools and clinical decision support, rather than reiterating the need for early treatment following detection.

Evidence

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

Evidence in the form of randomized controlled trials (RCTs), observational studies, and literature reviews were assessed to understand the frequency and dangers of delayed VTE diagnosis.

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

VTE is a commonly missed or delayed diagnosis (Schiff et al., 2009) as it is difficult for clinicians to diagnosis due to non-specific symptoms (Ageno et al., 2008;). Whalen et. al. estimated primary care diagnostic delay at 3.9 days with diagnostic delay defined as the number of days between symptom onset and the time of diagnosis (2016). However, there are no existing measures of VTE diagnostic delay in primary care or other settings to systematically quantify and routinely track this problem.

There is a stark contrast in mortality between patients who receive immediate diagnosis and treatment of VTE and those who are left undiagnosed (Liederman et al., 2020). VTEs are associated with a high 30-day mortality rate (Tagalakis et al., 2013), and delays in VTE diagnosis are associated with higher rates of complications and an increased risk of mortality (Klok et al., 2018).

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. Untreated pulmonary embolisms have a mortality rate of approximately 30%, and nearly 30% of untreated DVTs will result in severe swelling or ulceration of the leg (Raju et al., 1986; Benotti et al., 1984). With prompt diagnosis and treatment, PE or treatment-related mortality is less than 1% (Büller et al., 2012).

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

A literature review was conducted in collaboration with the Harvard Countway librarian at Harvard Medical School. Information was obtained by conducting applicable VTE searches through PubMed and utilizing the similar article feature until it no longer reflected the measure specifications. The literature review involves a compilation of journals relevant to the specifications and criteria for this eCQM. The focus of the review was to understand the diagnostic recommendations of VTE and the occurrence of DVT and PE in the entire adult population.

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

1. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, Elstein AS, Hasler S, Kabongo ML, Krosnjar N, Odwazny R, Wisniewski MF, McNutt RA. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med. 2009 Nov 9;169(20):1881-7. doi: 10.1001/archinternmed.2009.333. PMID: 19901140.
2. Ageno W, Agnelli G, Imberti D, Moia M, Palareti G, Pistelli R, Rossi R, Verso M, MASTER Investigators. Factors associated with the timing of diagnosis of venous thromboembolism: results from the MASTER registry. Thrombosis research. 2008 Jan 1;121(6):751-6.
3. Walen S, Damoiseaux RA, Uil SM, van den Berg JW. Diagnostic delay of pulmonary embolism in primary and secondary care: a retrospective cohort study. Br J Gen Pract. 2016;66(647):e444-50.
4. Liederman Z, Chan N, Bhagirath V. Current challenges in diagnosis of venous thromboembolism. Journal of Clinical Medicine. 2020 Oct 29;9(11):3509.
5. Tagalakis V, Patenaude V, Kahn SR, Suissa S. Incidence of and mortality from venous thromboembolism in a real-world population: the Q-VTE Study Cohort. Am J Med. 2013 Sep;126(9):832.e13-21. doi: 10.1016/j.amjmed.2013.02.024. Epub 2013 Jul 3. PMID: 23830539.
6. Klok FA, Barco S, Konstantinides SV, Dartevelle P, Fadel E, Jenkins D, Kim NH, Madani M, Matsubara H, Mayer E, Pepke-Zaba J, Delcroix M, Lang IM. Determinants of diagnostic delay in chronic thromboembolic pulmonary hypertension: results from the European CTEPH Registry. Eur Respir J. 2018 Dec 6;52(6):1801687. doi: 10.1183/13993003.01687-2018. PMID: 30409820.
7. Dalen JE. Pulmonary embolism: what have we learned since virchow?: natural history, pathophysiology, and diagnosis. Chest. 2002 Oct 1;122(4):1440-56.
8. 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.
9. Raju S, Fredericks RK. Late hemodynamic sequelae of deep venous thrombosis. Journal of vascular surgery. 1986 Jul 1;4(1):73-9.
10. Benotti JR, Dalen JE. The natural history of pulmonary embolism. Clinics in chest medicine. 1984 Sep 1;5(3):403-10.
11. Büller HR, Prins MH, Lensing AW, Decousus H, Jacobson BF, Minar E, Chlumsky J, Verhamme P, Wells P, Agnelli G, Cohen A. Oral rivaroxaban for the treatment of symptomatic pulmonary embolism. The New England journal of medicine. 2012;366(14):1287-97.

Importance to Measure and Report: Gap in Care/Disparities (1b.01 - 1b.05)

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.

VTE is a serious, preventable public health problem affecting approximately 300,000–600,000 individuals in the U.S. each year and requires timely and adequate treatment (Beckman, et al. 2010). VTE consists of pulmonary embolism and deep vein thrombosis and its 30-day mortality rate is up to 23% (Tagalakis, et al. 2013; Nijekuter, et al. 2007). Because signs and symptoms of VTE are non-specific, timely recognition of VTE is difficult, and missed VTE diagnosis is common. Two classic studies of necropsies in large hospitals found that 9%-12% had VTE and 84%-91% were undiagnosed at the time of death (Karwinksi, et al. 1989, Carvalho Bricola, et al. 2013).

In addition to concerns over patient safety, VTE events are costly to healthcare systems. Ruppert et al. (2011) estimated that VTE complications ranged from $426-$41,133 across literature and represent a financial burden on healthcare systems. Preventing VTE events prevents resulting adverse events from occurring, meaning that this eCQM has the potential to save thousands of dollars in avoided healthcare costs at the patient level.

In 2008, the U.S. Surgeon General declared VTE a public health emergency and issued an official call for action to prevent DVT and PE. The surgeon general warned that while morbidity and mortality related to other deadly cardiovascular diseases have greatly improved over the past decade, VTE-related outcomes have not improved and without extensive efforts this problem will worsen as the population ages (DHHS, 2008).

In 2019, the American Society of Hematology published VTE diagnosis guidelines to provide an evidence-based strategy to efficiently evaluate patients (Anderson et al., 2019). The goal of these guidelines is to improve diagnostic accuracy by assisting providers with evaluating patients with suspected VTE while reducing unnecessary and more invasive testing (Lim, et al. 2018). While routine use of guidelines in primary care would likely reduce the number of missed or delayed VTE diagnoses, integration into practice is challenging as VTE symptoms are nonspecific and often present as symptoms consistent with an underlying chronic illness.

Strategies such as measurement of diagnostic performance are needed to assist primary care providers with adopting VTE diagnosis guidelines and routinely using them in clinical practice. Currently there is no way to measure VTE diagnostic performance. Metrics are needed to quantify suboptimal VTE diagnostic performance, improve early recognition of VTE symptoms, and ultimately reduce unfavorable VTE outcomes.

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 eCQM. 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. This measure has the potential to lower health care costs associated with VTE by providing ongoing patient outcome data that can be used to improve VTE diagnostic performance and to reduce complications associated with delayed diagnosis and treatment.

Citations:

1. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous Thromboembolism: A Public Health Concern. Am J Prev Med. 2010 Apr 1;38(4, Supplement):S495–501.

2. Tagalakis V, Patenaude V, Kahn SR, Suissa S. Incidence of and Mortality from Venous Thromboembolism in a Real-world Population: The Q-VTE Study Cohort. Am J Med. 2013 Sep 1;126(9):832.e13-832.e21.
3. Nijkeuter, M., Söhne, M., Tick, L.W., Kamphuisen, P.W., Kramer, M.H., Laterveer, L., van Houten, A.A., Kruip, M.J., Leebeek, F.W., Büller, H.R. and Huisman, M.V., 2007. The natural course of hemodynamically stable pulmonary embolism: clinical outcome and risk factors in a large prospective cohort study. *Chest*, *131*(2), pp.517-523.
4. Karwinski B, Svendsen E. Comparison of clinical and postmortem diagnosis of pulmonary embolism. J Clin Pathol. 1989 Feb;42(2):135-9. doi: 10.1136/jcp.42.2.135. PMID: 2921354; PMCID: PMC1141815.
5. Carvalho Bricola, S.A.P., Paiva, E.F., Lichtenstein, A., Gianini, R.J., Duarte, J.G., Shinjo, S.K., Eluf-Neto, J. and Arruda Martins, M., 2013. Fatal pulmonary embolism in hospitalized patients: a large autopsy-based matched case-control study. *Clinics*, *68*, pp.679-685.
6. Ruppert A, Steinle T, Lees M. Economic burden of venous thromboembolism: a systematic review. Journal of medical economics. 2011 Jan 1;14(1):65-74.
7. *The Surgeon General's call to action to prevent deep vein thrombosis and pulmonary embolism.* United States Department of Health and Human Services. Office of the Surgeon General (US) CTI - Publications and Reports of the Surgeon General; 2008.
8. Anderson, D.R., Morgano, G.P., Bennett, C., Dentali, F., Francis, C.W., Garcia, D.A., Kahn, S.R., Rahman, M., Rajasekhar, A., Rogers, F.B. and Smythe, M.A., 2019. American Society of Hematology 2019 guidelines for management of venous thromboembolism: prevention of venous thromboembolism in surgical hospitalized patients. *Blood advances*, *3*(23), pp.3898-3944.
9. Lim, W., Le Gal, G., Bates, S.M., Righini, M., Haramati, L.B., Lang, E., Kline, J.A., Chasteen, S., Snyder, M., Patel, P. and Bhatt, M., 2018. American Society of Hematology 2018 guidelines for management of venous thromboembolism: diagnosis of venous thromboembolism. *Blood advances*, *2*(22), pp.3226-3256.

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.

Site 1: Boston, MA area
Sample size (total denominator): 3,591
Mean performance score: 72.6%
Standard deviation: 21.52%
Min performance score: 0%
Max performance score: 100%
Interquartile range for performance scores: 40%
Performance scores by decile:

Decile	DOVE Rate (%)
10	33.33
20	50
30	64.17

40	66.67
50	74.34
60	84.62
70	100
80	100
90	100
100	100

Data source: A total of 214 primary care clinician groups across the Site 1 healthcare system were included in the analysis. 5,514 patient encounters were assessed, 3,591 of which met the inclusion criteria for the measure. Descriptive statistics of patients included and excluded in the measure can be found in the reliability testing section.
Dates of data: 01/06/2016 - 12/31/2021
**Clinician groups with ≥1 encounter were included in performance score data **

Site 2: Lexington, KY area
Sample size (total denominator): 245
Mean performance score: 77.14%
Standard deviation: N/A, rate calculated at the organization level (single entity)
Min performance score: 77.14%
Max performance score: 77.14%
Interquartile range for performance scores: N/A, rate calculated at the organization level
Performance scores by decile: N/A, rate calculated at the organization level
Data source: Site 2 was assessed at the organization-level as this healthcare system did not have an interoperable EHR at the time of analysis. 632 patient encounters were assessed, of which 245 patients met the measure inclusion criteria. This large dropout rate in Site 2 is attributable to the lack of interoperability in the site EHR, and reflects our measure specifications for this eCQM to be utilized in interoperable healthcare systems. Descriptive statistics of patients included and excluded in the measure can be found in the reliability testing section.
Dates of data: 12/01/2016 - 12/31/2020

Site 3: Hershey, PA area: data collection and rate calculation currently in progress.

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.

Not applicable, performance data on the measure are available.

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.

A sub analysis was performed to assess disparities by social determinant of health variables in Site 1 patients (**Table 3**). A T-test assuming equal variances was performed to assess if there was significant variation between the subsamples. No significant differences were found across patients by race, ethnicity, sex, insurance, and age (p values ranged from 0.387-0.980 by variable), meaning there were no significant differences in delayed VTE diagnosis rate by patient characteristic. This sub analysis reenforces the rationale to not perform risk adjustment or stratification on this eCQM as no significant differences were identified between patient demographics. As discussed in section 1b.05, this measure only includes VTE encounters where a patient has reported VTE symptoms to a primary care provider within 30 days, meaning a provider had information available to assess for the presence of a VTE.

Table 3: Sub Analysis of Site 1 Patients by Race, Ethnicity, Sex, Insurance, and Age

Demographics	**Denominator (column %)**	**Numerator (column %)**	**Dove Rate**	**P Value**
Total sample:	3591	2607	72.60%	
Race:				
Black:	312 (8.69)	213 (8.17)	68.27%	**0.841**
White:	2945 (82.01)	2142 (82.16)	72.73%	
Other:*	334 (9.30)	252 (9.67)	75.45%	
Ethnicity:				
Hispanic:	233 (6.49)	172 (6.60)	73.82%	**0.548**
non-Hispanic:	3290 (91.62)	2389 (91.64)	72.61%	
Missing/declined:	68 (1.89)	46 (1.76)	67.65%	
Sex:				
Female:	1847 (51.43)	1346 (51.63)	72.87%	**0.979**
Male:	1744 (48.57)	1261 (48.37)	72.31%	
Insurance:				
Public:	1969 (54.83)	1472 (56.46)	74.76%	**0.387**
Private:	1611 (44.86)	1128 (43.27)	70.02%	
Other:**	11 (0.31)	7 (0.27)	63.64%	
Age:				
<65:	1509 (42.02)	1047 (40.16)	69.38%	**0.888**
≥65:	2082 (57.98)	1560 (59.84)	74.93%	

*Other racial category includes Asian, American Indian, Alaska Native, and race self-reported as "other"
**Other insurance category includes free care from the hospital and self-pay

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.

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 disparities in VTE prevalence by African American race are apparent across published literature, this eCQM specifically targets delayed diagnosis of VTE after symptoms are recorded in a primary care visit. As only cases where a provider has information available to make a diagnosis are included in the measure, this eCQM is not risk adjusted or stratified for any of the disparity variables highlighted above. As seen in the sub analysis in section 1b.04, delayed VTE diagnosis rates are not significantly different by race, ethnicity, sex, insurance type, or age.

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.

Additionally, this eCQM is designed to work alongside a separate but related CDS tool to provide recommendations to clinicians for patients who are recognized as at higher risk for a VTE event, which does not involve a risk-adjustment component. 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 CDS predictors than with the inclusion of a risk adjustment model.

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 HBGoldberg 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 MDHeit JAMohr DNPetterson TMO'Fallon WMMelton 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.

Scientific Acceptability: Maintenance (2ma.01 - 2ma.04)

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

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

Current Submission:
Updated testing information here.

Previous Submission:
Testing from the previous submission here.

☐ Yes
☐ No

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

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

Current Submission:
Updated testing information here.
Previous Submission:
Testing from the previous submission here.

☐ Yes
☐ No

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

☐ Yes
☐ No

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

social risk factors.

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

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

☐ Yes - Additional risk adjustment analysis is included

☐ No additional risk adjustment analysis included

Scientific Acceptability: Reliability - Testing (2a.01 - 2a.12)

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 Battelle staff at PQMsupport@battelle.org 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 Battelle staff at PQMsupport@battelle.org with any questions.
- 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 the 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

OR

• 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.

☐ Assessment Data
☐ Claims
☐ Electronic Health Data
☒ Electronic Health Records
☐ Instrument-Based Data
☐ Management Data
☐ Other (please specify here:)
☐ Paper Medical Records
☐ Registry Data

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).

N/A, existing dataset not used. For measure testing, data were pulled from two U.S. academic healthcare systems using Epic and Allscripts EHR. Testing in a third site with Cerner EHR is currently underway.

2a.03) Provide the dates of the data used in testing.
Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

Site 1: 01/06/2016 - 12/31/2021
Site 2: 12/01/2016 - 12/31/2020
Site 3: to be determined.

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.

☐ Accountable Care Organization
☒ Clinician: Group/Practice
☐ Clinician: Individual

☐ Facility
☐ Health Plan
☒ Integrated Delivery System
☐ Other (specify)
☐ Population: Community, County or City
☐ Population: Regional and State

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.

The Site 1 sample included a total of 214 primary care sites. 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 Site 2 encounters did not meet inclusion criteria, compared to 34.87% in Site 1). Accessing care across sites is a limitation of eCQMs in non-interoperable systems and is not limited to the DOVE eCQM. Based on testing in Site 2, we have determined that the measure would be most meaningful when used within an integrated care delivery network. We are currently testing the measure in a third site that is an integrated care delivery network.

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.

Table 4 displays the descriptive statistics of patients who met the inclusion criteria for the DOVE eCQM. **Table 5** 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.

Table 4: Descriptive Statistics of the Included Sample

	Site 1	**Site 2**
Number of total encounters	5,514	632
Encounters included in the measure (%)	3,591 (65.13)	245 (38.77)
Encounters excluded from the measure (%)	1,923 (34.87)	387 (61.23)
Number of delayed VTE diagnosis events	2,607	189
Site delayed diagnosis rate	72.60%	77.14%

Number of clinician groups	214	1
Age:		
Mean age at VTE (SD)	65.89 (15.27)	58.14 (N/A)
Age ≥65 (%)	2,082 (57.98)	84 (34.29)
Age <65 (%)	1,509 (42.02)	161 (65.71)
Sex (%):		
Female	1,847 (51.43)	129 (52.65)
Male	1,744 (48.57)	116 (47.35)
Self-Reported Race (%):		
Black/African American (%)	312 (8.73)	24 (9.80)
White (%)	2,945 (82.01)	221 (90.20)
Other* (%)	334 (9.30)	0 (0)
Self-Reported Ethnicity (%):		
Hispanic	233 (6.49)	4 (1.63)
Non-Hispanic	3,290 (91.62)	236 (96.33)
Missing/declined ethnicity	68 (1.89)	5 (2.04)
Insurance Type (%):		
Public Insurance	1,969 (54.83)	162 (66.12)
Private Insurance	1,611 (44.86)	83 (33.88)
Other insurance**	11 (0.31)	0 (0)
English as a first language (%)	3,325 (92.59)	241 (98.37)
Mean number of VTE symptoms (SD)	2.31 (1.34)	1.4 (N/A)
Median income (via ZIP Code) (SD)	$74,359 ($27,059)	$38,254 (N/A)

**Other category includes Asian, American Indian or Alaska Native, and race self-reported as "other"*
***Other insurance category includes self-pay and free care*

Table 5: Descriptive Statistics of the Excluded Sample

	Site 1	Site 2
Number of total encounters	5,514	632
Encounters excluded from the measure (%)	1,923 (34.87)	387 (61.23)
Number of clinician groups	170	1
Age:		
Mean age at VTE (SD)	63.86 (15.26)	55.61 (N/A)
Age ≥65 (%)	1,026 (53.35)	131 (33.85)
Age <65 (%)	897 (46.65)	256 (66.15)
Sex (%):		
Female	930 (48.36)	129 (52.65)
Male	993 (51.64)	116 (47.35)
Self-Reported Race (%):		
Black/African American	177 (9.20)	58 (14.99)
White	1,571 (81.70)	326 (84.24)
Other*	175 (9.10)	3 (0.78)
Self-Reported Ethnicity (%):		
Hispanic	140 (7.28)	9 (2.33)

Non-Hispanic	1,746 (90.80)	373 (96.38)
Missing/declined ethnicity	37 (1.92)	5 (1.29)
Insurance Type (%):		
Public Insurance	1,009 (52.47)	298 (77.0)
Private Insurance	911 (47.37)	85 (21.96)
Other insurance	3 (0.16)	4 (1.03)
English as a first language (%)	1,757 (91.37)	375 (96.90)
Median income (via ZIP Code) (SD)	73,823 (27,112)	$38,965 (N/A)

**Other category includes Asian, American Indian or Alaska Native, and race self-reported as "other"*
***Other insurance category includes free care and self-pay*

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.

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

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.
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.

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

2a.09) Select the level of reliability testing conducted.
Choose one or both levels.
☒ Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)
☒ Accountable Entity Level (e.g., signal-to-noise analysis)

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.

The following forms of reliability testing were conducted:

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

NLP Phenotyping Algorithm Accuracy Testing:

We developed a rule-based symptom extractor to identify VTE symptoms in primary care clinical notes. For evaluation, we used both a case cohort and a control cohort. The case cohort included patients who met our inclusion criteria based on ICD 10 codes, imaging CPT codes, and RxNorm codes. The control cohort included patients who did not meet the inclusion criteria.

We used a random sample of 279 case cohort 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. This approach was used to evaluate a sample of 50 control cohort patients. We measured precision (PPV), recall (sensitivity), specificity, and NPV with a total of 26 rounds for patients with a VTE diagnosis (case cohort), and 5 rounds for patients with no VTE diagnosis (control cohort).

Inter rater NLP reliability was assessed by sharing deidentified notes across Sites 1 and 2 to ensure consistency in NLP algorithm performance. The NLP algorithm was performed on clinical notes from 15 encounters in each site (30 total) and compared to ensure that symptom extractions from both sites were consistently similar. Agreement was based on the binary presence of zero or one or more symptoms.

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 3: Novel VTE Phenotype cohort development and analytic pipeline

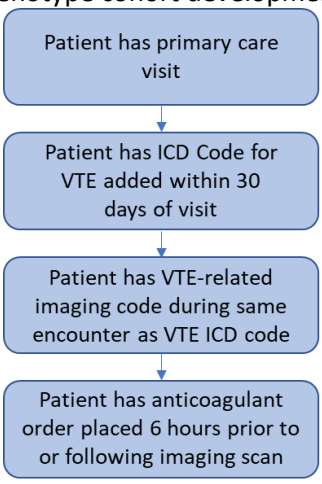

Chart Reviews and Algorithm Performance: we calculated the accuracy novel VTE phenotyping pipeline by calculating the positive predictive value (PPV), the negative predictive value (NPV), the sensitivity, and the specificity of chart reviews.

- 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 cart 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.

Signal to Noise Analysis:
A signal to noise analysis was conducted for the 15 largest practices at the individual clinician level (n=29) and the clinician group level (n=15) in Site 1. A signal to noise analysis estimates the proportion of overall variability explained by the differences between measured entities (between individual clinicians, between clinician groups). A minimum sample size of 10 encounters was required for the signal to noise analysis.

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, Measure Evaluation Criteria).

NLP Algorithm Accuracy Testing:
26 rounds of chart review were conducted with patients who had a VTE diagnosis (case cohort). 5 rounds of chart reviews were conducted with patients without VTEs (control cohort). 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 sweling 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*

Table 6: NLP Algorithm Accuracy Testing for Patients with a VTE Diagnosis (Abridged)

	Round 1 (n=673 note sentences)	Round 26 (n=938)
Precision (PPV) (95% CI)	0.50 (0.42-0.58)	1.00 (1.00-1.00)
Recall (sensitivity) (95% CI)	0.86 (0.74-0.94)	1.00 (0.92-1.00)
Specificity (95% CI)	0.93 (0.91-0.95)	1.00 (1.00-1.00)
NPV (95% CI)	0.99 (0.98-0.99)	1.00 (1.00-1.00)

Table 7: NLP Algorithm Accuracy Testing for Patients with no VTE Diagnosis (Abridged)

	Round 1 (n=281)	Round 5 (n=912)
Precision (PPV) (95% CI)	0.53 (0.36-0.70)	0.85 (0.64-0.95)
Recall (sensitivity) (95% CI)	1.00 (0.63-1.00)	0.90 (0.67-0.99)
Specificity (95% CI)	0.97 (0.95-0.99)	1.00 (0.99-1.00)
NPV (95% CI)	1.00 (1.00-1.00)	1.00 (0.99-1.00)

Inter rater reliability of the NLP algorithm was 100% across 30 clinical encounter notes in two sites. The final Kappa was 1.00, representing almost perfect agreement.

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 8 below.

Table 8: Value set codes used to indicate a VTE

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.

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 algorithms PPV was 95.80%.
Most of the false positives identified were instances where the provider suspected a pulmonary embolism (PE), 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.

Table 9: Chart Review Results

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

Signal to Noise Analysis:

At the clinician group level, 15 groups from Site 1 were sampled and the median signal-to-noise

statistical result was 0.5393 (95% CI; 0.4166, 0.6619). The minimum SNR was 0.37512 and the maximum was 0.9489.
At the clinician level, 29 groups from Site 1 were sampled and the median signal-to-noise statistical result was 0.19054 (95% CI: 0.1544, 0.1983). The minimum SNR was 0.1225, the maximum was 0.3262.

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?)

NLP Algorithm Accuracy Testing:
Inter rater reliability of the NLP algorithm demonstrated that our NLP algorithm can reliably extract VTE symptoms from clinical notes when used across healthcare systems. The final VTE phenotyping algorithm was successful in accurately and reliably identifying VTE cases from structured data in ICD-10, imaging, and RxNorm codes.
Unstructured data has previously been inaccessible in eCQMs, 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 eCQMs is particularly powerful for complicated disease conditions in large-scale patient populations, like VTE diagnosis in integrated healthcare systems.

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.

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 4**):

- DOVE eCQM methods (ICD-10 CM, imaging, and RxNorm codes) PPV = 95.8%
- Alotaibi et al., 2015 (ICD-10 and imaging) PPV = 73.3%
- Fang et al., 2017 methods (ICD-10 only) PPV = 64.6%

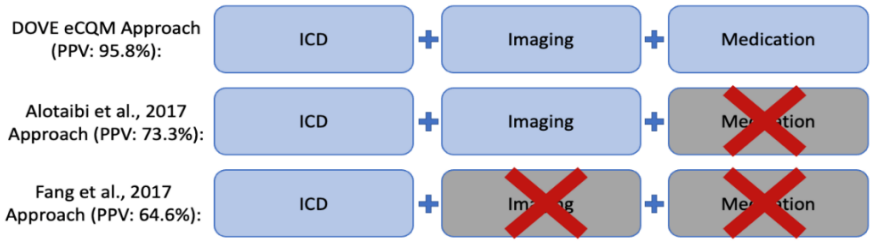

Figure 4: Alternative VTE Cohorts

Signal to Noise Analysis:

This measure is specified for use at the clinician group level; per MIPS-Quality requirements, both individual clinician level and clinician group level Signal to Noise analyses have been conducted. The SNR was stronger at the clinician group level compared to the clinician level, reenforcing our rationale as to why this measure is specified at the group level.

The results from the signal to noise analysis appear weak, but are strengthened by context; measures with less nuance (like two providers having the same findings from a review of a medication order) may have a higher SNR than our measure, but our measure assesses a very complex and frequently missed condition using strict criteria (self-reported VTE events in a primary care encounter prior to the diagnosis of a VTE characterized by co-occurring imaging, medication, and ICD-10 codes). Additionally, the sample sizes for this analysis at the clinician and clinician group level were small (15 and 29, respectively), which impacted the resulting SNR.

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.
4. Alotaibi GS, Wu C, Senthilselvan A, McMurtry MS. The validity of ICD codes coupled with imaging procedure codes for identifying acute venous thromboembolism using administrative data. Vasc Med. 2015 Aug;20(4):364-8. doi: 10.1177/1358863X15573839. Epub 2015 Apr 1. PMID: 25834115.
5. Fang MC, Fan D, Sung SH, Witt DM, Schmelzer JR, Steinhubl SR, Yale SH, Go AS. Validity of Using Inpatient and Outpatient Administrative Codes to Identify Acute Venous Thromboembolism: The CVRN VTE Study. Med Care. 2017 Dec;55(12):e137-e143. doi: 10.1097/MLR.0000000000000524. PMID: 29135777; PMCID: PMC5125903.

Scientific Acceptability: Validity - Testing (2b.01 - 2b.04)

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

☒ Patient or Encounter-Level (data element validity must address ALL critical data elements)
☒ Accountable Entity Level (e.g., hospitals, clinicians)
☐ Empirical validity testing of the measure score
☒ 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)

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.

Validity testing was assessed in the following ways:

- Chart reviews comparing manual review with the eCQM (patient or encounter-level testing)
- Face validity from a technical expert panel (TEP)
- Random split sample correlation and rate calculation (accountable entity level testing)

Manual Chart Review:

Manual chart reviews were performed on a random sample of 30 patients from Site 1 to compare to the eCQM. A research assistant was blinded to the VTE status of the patient (did not have a VTE, had a VTE with timely diagnosis, had a delayed diagnosis VTE event), reviewed the patient's chart, and manually identified if the patient would be excluded from the measure, meet the denominator criteria, or meet the numerator criteria. The purpose of these chart reviews was to determine the level of agreement between manual EHR review and the eCQM. Clinicians from Site 2 are currently in the process of performing chart reviews.

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.

During a July 2022 meeting, the TEP was presented final measure specifications, initial rate calculations, and information on delayed diagnosis of VTE across literature. The TEP also had an opportunity to discuss questions and provide feedback to the measure development team. A formal face validity vote was conducted via an online survey (Google Poll) that was sent to the TEP members by email after the presentation and discussion. Only TEP members who were present for this meeting were eligible to participate in the face validity vote. The survey asked the following face validity question: "**The VTE Diagnostic Delay in Primary Care eCQM, as specified, can be used to distinguish good form poor clinician group-level quality related to patient safety"**. TEP members were blinded to each other's

responses, but were told the final face validity vote after all eligible members had voted.

Random Half Split Correlation and Rate Calculation

A random half split correlation was conducted at the clinician group level in Site 1, with 15 clinician groups included in the analysis. To perform a random half split correlation analysis, we required a minimum of 20 encounters for each eligible clinician (10 encounters in each split sample). Encounters in each clinician group were randomly split into a test group or a validation group, with 50% of encounters in each group. The descriptive statistics and p-values for each group were calculated. A Spearman correlation and ICC with 95% confidence intervals were calculated. The ICC was calculated to describe how much variation in the provider-group level scores is due to provider-group level signal variation. The spearman correlation coefficient was calculated to compare the relative rankings of clinician groups in the test and validation samples.

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

Examples may include correlations or t-test results.

Manual Chart Review:

Following the manual chart review of 30 patients from Site 1, 22 patients were sorted into the denominator, 9 of the 22 patients from the denominator were included in the numerator, and 8 patients were excluded from the measure. Manual chart review and the eCQM had 100% agreement (kappa = 1.0, PPV=100%, NPV=100%), demonstrating strong validity and agreement in the eCQM.

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 form poor clinician group-level quality related to patient safety."** The final vote was 5/5 (100%) in agreement with the voting statement among present members. 1 member was absent and did not vote.

Random Half Split Correlation and Rate Calculation

15 clinician groups were included in the analyses from Site 1. 1,168 encounters from 15 clinician groups were included in the test sample, 1,177 encounters from 15 clinician groups were included in the validation sample. The DOVE rate in the test sample was 72.52%, the DOVE rate in the validation sample was 75.02%. P values were calculated for encounter-level demographics, no variables were significantly different between test and validation groups.

The Spearman correlated was 0.7817 (95% CI: 0.4372, 0.9429). The ICC in the test sample was 0.0174 (95% CI: 0.0042, 0.6701). The ICC in the validation sample was 0.0262 (95% CI: 0.0086, 0.2654).

Table 10: Descriptive Statistics of the Test and Validation Samples

	Site 1 Test Sample	Site 1 Validation Sample	P Value
Number of encounters, n	1168	1177	
Encounters included in the measure, n	1168	1177	
Encounters excluded from the measure, n	0	0	
Number of delayed VTE diagnosis events, n	847	883	
Site delayed diagnosis rate, %	72.52	75.02	
Number of clinician groups, n	15	15	

Age:			
Mean age at VTE (SD)	66.49 (14.58)	65.39 (15.05)	0.16
Age ≥65 (%)	96 (59.59)	659 (55.99)	
Age <65 (%)	472 (40.41)	518 (44.01)	
Sex (%):			
Female	608 (52.05)	620 (52.68)	0.73
Male	560 (47.95)	557 (47.32)	
Self-Reported Race (%):			
Black/African American	131 (11.22)	110 (9.35)	0.95
White	936 (80.14)	950 (80.71)	
Other*	101 (8.65)	117 (9.94)	
Self-Reported Ethnicity (%):			
Hispanic	62 (5.31)	77 (6.54)	0.47
Non-Hispanic	1,076 (92.12)	1,076 (91.42)	
Missing/Declined	30 (2.57)	24 (2.03)	
Insurance Type (%):			
Public Insurance	645 (55.22)	638 (54.20)	0.94
Private Insurance	523 (44.44)	539 (45.63)	
Other insurance**	4 (0.34)	2 (0.17)	
English as a first language (%)	1,096 (93.84)	1,075 (91.33)	0.24
Median income (via ZIP Code), USD (SD)	73,800 (27,360)	73,610 (27,620)	0.46

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?)

Manual Chart Review:

Manual chart reviews compared against the eCQM were performed to evaluate the validity of the eCQM in identifying numerator, denominator, and exclusion encounters. The kappa, PPV, and NPV for the manual chart review and eCQM comparison was 100%, demonstrating that this eCQM can accurately define eligible VTE events.

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.

Random Half Split Correlation and Rate Calculation

Spearman's rank correlation computed to assess the ranking of DOVE rates between the test and validation samples showed a **strong positive** correlation between the two samples (0.78175). An ICC was calculated in the complete sample to describe how much variation in the provider-group level scores is due to provider-group level signal variation; the ICC was low (0.0174 in the test sample, 0.02616 in the validation sample).

Scientific Acceptability: Validity - Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) (2b.05 - 2b.14)

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.

Subgroup Analysis: 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 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.

Benchmarking: To assess the performance gap for the DOVE eCQM, we used the Achievable Benchmarks of Care (ABC method) by Kiefe et al. (2001). The rationale for using this approach is that the ABC method takes into consideration groups with low numbers of eligible cases by adjusted for low denominators. The ABC approach to benchmarking has been shown to increase the effectiveness of clinician performance feedback (Kiefe et al., 2001). Complete methods for how to calculate the ABC method benchmark can be found in the cited article by Kiefe et al., 2001.
Benchmarking was assessed using the largest 15 clinician groups in Site 1. Each clinician group was randomly split into two samples. An average of 50% of each group was included in the clinician group sample in the Test sample, and an average of 50% of each group was included in the clinician group sample in the Validation sample.
An adjusted performance factor (APF) for the test and validation samples of all clinician groups was calculated. The APF can be calculated using the following formula: APF = (numerator +1)/(denominator +1). Test and validation clinician groups were then separately ranked from lowest to highest APF. Groups were ranked from low to high because a lower DOVE rate is indicative of higher quality care. The top performing clinician groups that accounted for a minimum of 10% of the overall population in the test and validation samples were identified as the benchmark group. The ABC Benchmark is then calculated using the sum of all and the sum of all denominators in the benchmark group, calculated separately for the test and validation samples using the following formula: Benchmark = (Σ numerator)/(Σ denominator).

References:

1. Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. Jama. 2001 Jun 13;285(22):2871-9.

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.

Subgroup Analysis: In the subgroup analysis, rates between cohorts ranged from 65.78%-77.14%, and variation within cohorts is seen in each cohort rate range (**Table 11**).

Table 11: Sub analysis by clinician group sample size

Clinician Group Level Cohort	Cohort sample size	# Of groups in the cohort	Total denominator count	Total numerator count	Cohort rate (%) (SD)	Rate range (%)
Site 1A	>100	4	1,755	1,316	74.99% (3.96%)	72.18%-81.51%
Site 1B	50-99	5	335	226	67.46% (10.54%)	54.55%-79.31%
Site 1C	25-49	19	643	458	71.23% (10.42%)	46.15%-86.49%
Site 1D	<25	22	374	246	65.78% (13.09%)	46.15%-100%
Site 2	Facility level	N/A	245	189	77.14 (N/A)	N/A

Benchmarking: 2,335 encounters were included in the Test sample, 2,354 encounters were included in the Validation sample across 15 clinician groups from Site 1. The top performing clinician groups that accounted for a minimum of 10% of the overall population (n=359) in the test and validation samples were identified as the benchmark group (**Table 12, 13**). In the test sample, 716 encounters from 8 clinician groups were included in the benchmark group. In the validation sample, 620 encounters from 4 clinician groups were included in the benchmark group. The sample sizes in these groups are much larger than the 10% threshold for the total population (n=395) because Test Site 1H and Validation Site 1D are top performing sites with a higher frequency of DOVE events than the minimum threshold.

Table 12: Test Sample Benchmark Rate (Site 1)

Location	Denominator	Numerator	Unadjusted DOVE Rate	APF Adjusted DOVE Rate
Test 1A	44	22	50.00%	51.11%
Test 1B	32	19	59.38%	60.61%
Test 1C	20	12	60.00%	61.90%
Test 1D	21	13	61.90%	63.64%
Test 1E	34	23	67.65%	68.57%
Test 1F	22	15	68.18%	69.57%
Test 1G	20	14	70.00%	71.43%
Test 1H	523	383	73.23%	73.28%
Sum	**716**	**501**	**69.97%**	**70.01%**

Table 13: Validation Sample Benchmark Rate (Site 1)

Location	Denominator	Numerator	Unadjusted DOVE Rate	APF Adjusted DOVE Rate
Validation 1A	21	12	57.14%	59.09%
Validation 1B	44	26	59.09%	60%
Validation 1C	32	21	65.63%	66.66%
Validation 1D	523	372	71.13%	71.18%
Sum:	**620**	**431**	**69.52%**	**69.57%**

Using the benchmark formula, the benchmark for the test sample is 69.97%, the benchmark for the validation sample is 69.52%. The mean benchmark in Site 1 is 67.74%.

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?

In the subgroup analysis, high rates with some variation were identified in each cohort. This demonstrates that the DOVE eCQM 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 practice size. The variation in rates points to opportunities for quality improvement at the clinician group level.

Considering the mean DOVE rate across sites is upwards of 80% in many clinician groups, reaching the benchmark for poorly performing clinician groups would result in a clinically significant reduction in cases of delayed diagnosis of VTE.

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.

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.

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).

Table 14: Frequency of Data Elements

	Site 1 Available (%)	Site 1 Missing (%)	Site 2 Available (%)	Site 2 Missing (%)
Total eligible encounters:	3591	N/A	245	N/A
Age at VTE	3591 (100)	0 (0)	245 (100)	0 (0)
Sex	3591 (100)	0 (0)	245 (100)	0 (0)
Race	3591 (100)	0 (0)	245 (100)	0 (0)
Ethnicity	3536 (98.47)	55 (1.53)	240 (98.98)	5 (1.02)
Insurance type	3591 (100)	0 (0)	245 (100)	0 (0)
Language	3591 (100)	0 (0)	245 (100)	0 (0)
≥1 VTE symptom*	3591 (100)	0 (0)	245 (100)	0 (0)
Primary care encounter*	3591 (100)	0 (0)	245 (100)	0 (0)
VTE imaging scan*	3591 (100)	0 (0)	245 (100)	0 (0)
RxNorm anticoagulant order*	3591 (100)	0 (0)	245 (100)	0 (0)
VTE-related ICD billing codes*	3591 (100)	0 (0)	245 (100)	0 (0)
**required for measure calculation*				

As seen in **Table 14**, 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 (<2%), 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.

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.

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

specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

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 Site 2, a large nonintegrated care delivery network, that most patients reside in rural and nonmetro areas. As a result of location and the lack of EHR interoperability, many patients were excluded from the analysis as they received primary care and the VTE diagnosis in different locations. We recommend that the measure is implemented in an integrated care delivery network. This measure is not risk-adjusted and only has one set of specifications.

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

☐ Yes, there is more than one set of specifications for this measure
☒ No, there is only one set of specifications for this measure

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.

N/A, only one set of specifications.

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.

N/A, only one set of specifications.

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.

N/A, only one set of specifications.

Scientific Acceptability: Validity - Other Threats to Validity (Exclusions, Risk Adjustment) (2b.15 - 2b.32)

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

☐ N/A or no exclusions

☒ Yes, the measure uses exclusions.

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?

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.

In the Site 1 sample, 302 encounters were removed for hospice or palliative care within 90 days of the eligible encounter. In the Site 2 sample, 210 encounters were removed for hospice or palliative care within 90 days of the eligible encounter (**Table 15**).

Table 15: Encounters Before and After Applying Exclusions for Hospice and Palliative Care Encounters

	Sample Before Exclusion	Sample After Exclusion	% of Sample Lost
Site 1	3,893	3,591	0.078%
Site 2	842	632	24.94%

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.

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.

N/A

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.

N/A

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

☐ Statistical risk model with risk factors (specify number of risk factors)
☐ Stratification by risk category (specify number of categories)
☐ Other (please specify here:)
☒ No risk adjustment or stratification

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.

N/A, no statistical risk models used.

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.

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 drop out 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 not a 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 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 HBGoldberg 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.

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

☐ Published literature
☐ Internal data analysis
☒ Other (please specify here:)

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).

N/A, no statistical risk models or stratification used.

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.

N/A, no statistical risk models or stratification used.

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.

N/A, no statistical risk models or stratification used.

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.*

N/A, no statistical risk models or stratification used.

2b.27) Provide risk model discrimination statistics.
For example, provide c-statistics or R-squared values.

N/A, no statistical risk models or stratification used.

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

N/A, no statistical risk models or stratification used.

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.

N/A, no statistical risk models or stratification used.

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

N/A, no statistical risk models or stratification used.

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?

N/A, no statistical risk models or stratification used.

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.

N/A, no statistical risk models or stratification used.

Feasibility (3.01 - 3.07)

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

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

☐ Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

☐ Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

☐ Other (Please describe)

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. ALL data elements are in defined fields in electronic health records (EHRs)

☒ ALL data elements are in defined fields in electronic claims

☐ ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

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

☐ Some data elements are in defined fields in electronic sources

☐ No data elements are in defined fields in electronic sources

☐ Patient/family reported information (may be electronic or paper)

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.

N/A, all data elements needed to compute the performance measure score are from electronic sources.

3.04) Describe any efforts to develop an eCQM.

N/A, this measure is an eCQM.

3.05) Complete and attach the eCQM-Feasibility-Scorecard.xls file.

See attached NQF Feasibility Scorecard.

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.

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

This is a new measure and has not been implemented yet. Two U.S. healthcare systems were used in measure testing due to the time, financial, and logistic burdens associated with measure testing in multiple sites. Future analyses in additional healthcare systems are needed, testing in a third U.S. healthcare system is currently underway.

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.

N/A, there are no fees associated with the use of this measure. The NLP algorithm that can be used if the VTE symptoms are not available as structured data uses a no-fee license which includes an information overview on how the algorithm is used and instructions on implementation.

Use (4a.01 – 4a.10)

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.

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**

☐ Public Reporting
☐ Public Health/Disease Surveillance
☐ Payment Program
☐ Regulatory and Accreditation Programs
☐ Professional Certification or Recognition Program
☐ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☐ Quality Improvement (Internal to the specific organization)
☒ Not in use
☐ Use unknown
☐ Other (please specify here:)

4a.02) Check all planned uses.

☒ Public reporting
☐ Public Health/Disease Surveillance
☒ Payment Program
☐ Regulatory and Accreditation Program
☐ Professional Certification or Recognition Program
☒ Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
☒ Quality Improvement (internal to the specific organization)
☐ Measure Currently in Use
☐ Other (please specify here:)

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?

This is a new measure and has not yet been implemented for use.

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.

Specific program: Centers for Medicare & Medicaid Services, Merit-based Incentive Payment System (MIPS)
Purpose: The purpose of this measure is to quantify the rate of delayed diagnosis of VTE events after a patient has reported VTE symptoms to a provider in the primary care setting.
Intended Audience: This measure is designed for use by primary care clinician groups in integrated healthcare systems.
Timeline for implementing the measure within the specific timeframes: This measure will be submitted in May 2023 for potential inclusion as a Merit-based Incentive Payment System (MIPS) measure for the CMS Quality Payment Program (QPP). The measure is pending review by the Measure Applications Partnership if the measure is included on the MUC list. If the MAP recommends the measure for implementation and CMS approves, the measure will be included in the Physician Fee Schedule Proposed and Final Rule in 2024 for implementation in the QPP in 2025.
A plan for accountability applications addresses mechanisms for data aggregation and reporting: If the measure is implemented in the CMS Quality Payment Program through the rule-making process in summer of 2024, clinician-groups desiring to report on the measure for MIPS participation would begin collecting data for the measurement period starting January 1, 2025 and report measure performance data to CMS according to the eCQM data submission requirements described in the 2025 Physician Fee Schedule Final Rule, expected to be released in November of 2024.

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.

We performed several forms of stakeholder engagement throughout the measure development and testing process, including feedback from a technical expert panel (TEP) (described in validity testing), focus groups with providers and patients who had survived a VTE event, and public comment solicitation.

TEP: As part of the measure development process, the BWH team received feedback and guidance from our TEP comprised of six industry experts (**Table 16**). TEP members were identified by project investigators based on their experience and expertise relating to VTE.

Table 16: Names and Affiliations of DOVE TEP Members

Name	Institution	TEP Expertise/perspective
Jason Adelman	Columbia University Medical Center	Clinician
David W. Bates	Brigham and Women's Hospital	Clinician
Gregory Piazza	Brigham and Women's Hospital	Clinician
Isbelia Briceno	Oracle Cerner (EHR Vendor)	EHR expert
Martie Carnie	Brigham and Women's Hospital	Patient perspective
Tania Powell	Brigham and Women's Hospital	Patient perspective

Provider Interviews: Five providers participated in qualitative interviews during the environmental scan and measure development.

Patient Interviews: Five patients participated in qualitative interviews, all of whom had experienced a prior VTE.

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.

Technical Expert Panel: The eCQM development team met with the TEP and discussed the DOVE eCQM on 2 occasions throughout measure development (January, 2021, and July, 2022). The objective of these meetings was to determine the specifications that would allow the proposed measure to provide meaningful information that could be used to quantify delayed VTE diagnosis.

Provider Interviews: Interviews were between July and November 2021; all providers were identified by study co-investigators and TEP members. All provider interviews were performed at the individual level.

Patient Interviews: Four of the patients participated in a focus group in August 2022, and one patient participated in an interview in July 2021. Participants were identified by study co-investigators and TEP members.

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

TEP: In the January 2021 meeting, the TEP reviewed the preliminary measure specifications and recommended that the team move forward with measure development. In the July 2022 meeting, the TEP suggested performing a subgroup analysis based on patient characteristics to assess meaningful differences in by characteristic (see section 1b.04 for completed subgroup analysis), and provided the development team with contacts for patient focus groups. The TEP recommended that the team move forward with additional testing and a face validity vote was conducted. See validity testing for information on the face validity vote. Final face validity vote results were 100% in favor of the measure.

Provider Interviews:

Five providers participated in qualitative interviews during the environmental scan and measure development. Interviews were between July and November 2021; all providers were identified by study co-investigators and TEP members.

Measure Definition and Timeframe: Four out of five providers agreed with the proposed 24-hour timeline for a delayed diagnosis event following a primary care visit, two of whom said that they would also support a 24-36 hour timeline.

Potential Symptoms of VTE: Providers described a total of 16 potential symptoms of DVT and PE, six of which have been included in the measure (swelling of the leg and calf, redness of the leg and calf, pain of the leg and calf, chest pain, shortness of breath, lightheadedness, and syncope) based on consensus among providers, literature, and TEP feedback.

Measure Meaningfulness and Quality Improvement: All providers were interested in using this measure to drive quality improvement for reasons including holding providers and organizations accountable through public reporting, creating awareness about delayed diagnosis, providing audits and feedback to change physician behavior, and understanding what factors led to the diagnostic delay.

Patient Interviews:

Five patients participated in qualitative interviews, all of whom had experienced a prior VTE. Four of the patients participated in a focus group in August 2022, and one patient participated in an interview in July 2021. Participants were identified by study co-investigators and TEP members.

Encounters with Primary Care Providers Prior to the VTE diagnosis: Three of the five patients had direct contact with their primary care provider when VTE symptoms were present prior to their diagnosis, all of whom disclosed symptoms to their PCP. None of the three patients were referred for imaging or the ED following their PCP visit, signaling a potential delayed diagnoses.

Measure Meaningfulness and Quality Improvement: Patients shared that clinician group performance on this measure was unlikely to influence their selection of primary care providers, but would be meaningful to providers for quality improvement purposes.

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

N/A, all stakeholder engagement described above.

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

N/A, all stakeholder engagement described above.

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.

We used the feedback from our TEP to refine specifications like defining VTE by the presence of ICD-10, RxNorm, and CPT codes, and adding additional symptoms of VTE in the primary care setting for the NLP algorithm. Additional VTE symptoms were added to our symptom list based on feedback from focus groups with providers. The TEP and provider interviews were used to validate the measure specifications

from a clinician and EHR vendor perspective.

Usability (4b.01 - 4b.03)

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.

This is a new measure and there is no information available on performance improvement. This measure is not currently used in a program, but a primary goal of the measure is to provide information necessary for public reporting and quality improvement.

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

There were no identified unexpected findings during testing and the measure is not currently in use.

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

N/A, measure has not yet been implemented.

Related and Competing (5.01 - 5.06)

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 endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01) Search and select all endorsed related measures (conceptually, either same measure focus or target population) by going to the PQM website.
(Can search and select measures.)

N/A, no related measures.

5.02) Search and select all endorsed competing measures (conceptually, the measures have both the same measure focus or target population) by going to the PQM website.
(Can search and select measures.)

N/A, no competing measures.

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

N/A, no non-NQF endorsed related or competing measures.

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

☐ Yes
☐ No
N/A

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

N/A, no related or competing measures identified.

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.

N/A, no related or competing measures identified.

Additional (1 - 9)

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.

☐ Available in attached file

☒ No appendix

☐ Available at measure-specific web page URL identified in sp.09

2) List the workgroup/panel members' names and organizations.
Describe the members' role in measure development.

Table 17: Workgroup/Panel members' names and organizations

Name	Organization	Role in Measure Development
Jason Adelman	Columbia University Medical Center	TEP Member, clinician
David W. Bates, MD, MSc	Brigham and Women's Hospital	TEP Member, clinician
Gregory Piazza	Brigham and Women's Hospital	TEP Member, clinician
Isbelia Briceno	Cerner (EHR Vendor)	TEP Member, EHR expert
Martie Carnie	Brigham and Women's Hospital	TEP Member, Patient perspective
Tania Powell	Brigham and Women's Hospital	TEP Member, Patient perspective

3) Indicate the year the measure was first released.
N/A, new measure.

4) Indicate the month and year of the most recent revision.
N/A, new measure.

5) Indicate the frequency of review, or an update schedule, for this measure.
N/A, new measure.

6) Indicate the next scheduled update or review of this measure.
N/A, new measure.

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

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

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