Testing of an Electronic Clinical Quality Measure for Diagnostic Delay of Venous Thromboembolism (DOVE) in Primary Care

Patricia C. Dykes, PhD, RN[1,2]; Mica Bowen, BA[1]; Frank Chang, MSE[1]; Jin Chen, PhD[3]; Krissy Gray, AAS[3]; John Laurentiev, MS[1]; Luwei Liu, MS[1]; Purushottam Panta, MS[3]; Michael Sainlaire, MS[1]; Wenyu Song, PhD[1,2]; Ania Syrowatka, PhD[1,2]; Tien Thai, BS[1]; Li Zhou, MD, PhD[1,2]; David W. Bates, MD, MSc[1,2]; Lipika Samal, MD, MPH[1,2]; Stuart Lipsitz ScD[1,2]
[1]Brigham & Women's Hospital, Boston, MA; [2]Harvard Medical School, Boston, MA; [3]University of Kentucky, Lexington, KY
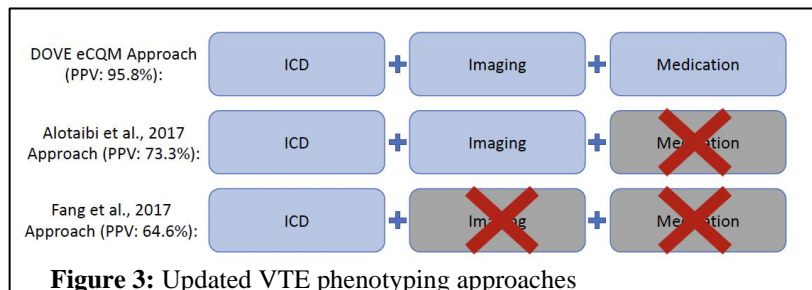
Abstract

Venous Thromboembolism (VTE) is a serious, preventable public health problem that requires timely treatment. Because signs and symptoms are non-specific, patients often present to primary care providers with VTE symptoms prior to diagnosis. Today there are no federal measurement tools in place to track delayed diagnosis of VTE. We developed and tested an electronic clinical quality measure (eCQM) to quantify Diagnostic Delay of Venous Thromboembolism (DOVE); the rate of avoidable delayed VTE events occurring in patients with a VTE who had reported VTE symptoms in primary care within 30 days of diagnosis. DOVE uses routinely collected EHR data without contributing to documentation burden. DOVE was tested in two geographically distant healthcare systems. Overall DOVE rates were 72.60% (site 1) and 77.14% (site 2). This novel, data-driven eCQM could inform healthcare providers and facilities about opportunities to improve care, strengthen incentives for quality improvement, and ultimately improve patient safety.

Introduction

Venous thromboembolism (VTE) includes pulmonary embolism and deep vein thrombosis. VTE is a commonly missed or delayed diagnosis(2, 3) affecting approximately 300,000-600,000 individuals in the U.S. each year, and requiring timely and adequate treatment to decrease mortality(4). The 30-day mortality rate is 23%(3, 5-7). Because signs and symptoms (s/s) of VTE are non-specific, timely recognition of VTE is difficult. 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(8, 9). Earlier diagnoses of VTE may reduce the morbidity and mortality associated with this dangerous condition and promote patient safety(10).

While patients often report symptoms to their primary care provider, VTE is a commonly missed or delayed diagnosis in primary care settings(11). One definition of diagnostic delay is the number of days between symptom onset and the time of diagnosis. Whalen et. al. estimated primary care diagnostic delay at 3.9 days(11). Despite the significant impact of diagnostic delay of VTE on patient outcomes, there is a notable absence of standardized measures to systematically quantify and routinely track this problem. With widespread adoption of electronic health records (EHR), data driven approaches for quality measurement are increasingly feasible. Electronic clinical quality measures or "eCQMs" are computerized tools that use EHR data to analyze and report on clinical performance, with the goal of improving patient outcomes and ensuring that healthcare services are safe, effective, patient-centered, timely, equitable, and efficient(12). Today a minority of measures used in the national quality payment program are eCQMs and none of the existing measures address VTE diagnostic delay(13). The lack of a standard data driven definition of VTE, as well as the low performance of existing identification algorithms points to a need for the novel DOVE eCQM.

Our team developed an eCQM that uses structured and unstructured EHR data to measure VTE diagnostic delay in primary care settings at the clinician group/practice and integrated delivery system levels. The DOVE eCQM is comprised of two algorithms that are described in detail elsewhere(14). First, based on the literature and advice from our TEP, we developed a multi-component VTE phenotyping algorithm that required a combination of ICD billing codes(15), CPT scan codes(16), and RxNorm anticoagulant treatment codes to identify patients with a VTE diagnosis. In comparison to existing VTE phenotyping algorithms using only ICD codes or a combination of ICD and CPT codes, this approach was superior in correctly identifying real VTE events and excluding events without VTE; our positive predictive value (PPV) was 95.8%(14), compared to 64.6% (ICD-10 code only)(15) and 73.3% (ICD-10 and CPT imaging codes)(16). In addition,

Figure 3: Updated VTE phenotyping approaches

we developed an NLP algorithm(17) to identify patient reported VTE signs and symptoms (s/s) documented in clinical notes during primary care visits to identify cases of incident VTE where primary care providers possessed relevant information potentially indicating VTE but did not refer patients for timely scanning or treatment, resulting in delayed diagnosis. Following NLP testing and refinement, the NLP algorithm had a PPV, negative predictive value (NPV), sensitivity, and specificity of 1.0 for patients with a VTE diagnosis, and a PPV of 0.85, NPV of 1.0, sensitivity of 0.9, and specificity of 1.0 for patients with no VTE diagnosis(18).

This manuscript describes the development and testing of the Diagnostic Delay of Venous Thromboembolism (DOVE) measure in two geographically distant healthcare systems, one urban/metropolitan and one rural, each using a different EHR system as its data source.

Methods

Alpha and beta testing were conducted using structured and unstructured EHR datasets. The project was reviewed and approved by the Mass General Brigham Human Subjects Committee and University of Kentucky IRB. The target population was adult patients aged 18 years and older diagnosed with VTE within 30 days of a primary care visit. EHR data were extracted from one large integrated care delivery network in the northeastern United States (U.S.) serving patients in urban and metropolitan areas from 01/06/2016 - 12/31/2021 and from a second large system in southern U.S. that serviced patients living in urban, metropolitan, and rural communities from 12/01/2016 - 12/31/2020. Descriptive statistics were calculated to characterize the sociodemographic information of the cohorts. Alpha testing was conducted to assess the feasibly of implementing the DOVE eCQM in an EHR system. Beta testing was conducted to assess data and measure reliability and validity and to calculate the DOVE rates across sites and practices within sites and to calculate a benchmark rate.

Development of measure specifications: To develop measure specifications, we conducted an environmental scan of available clinical guidelines, peer-reviewed and grey literature on VTE and common s/s. We also met with a technical expert panel (TEP) comprised of VTE experts (clinicians and researchers), measure development experts, and a patient who experienced a delayed VTE diagnosis after presenting in primary care with s/s. In addition, we conducted five individual interviews with providers and a focus group comprised of VTE survivors (n=5). Using this information, measure specifications were iteratively developed and refined through a multi-step optimization process. The final specifications for the DOVE eCQM that was tested included a denominator of all adult patients (aged 18 years and older) diagnosed with VTE who presented to primary care with VTE-related s/s within the 30 days before VTE diagnosis. The measure numerator is the subset of the denominator where the VTE diagnosis

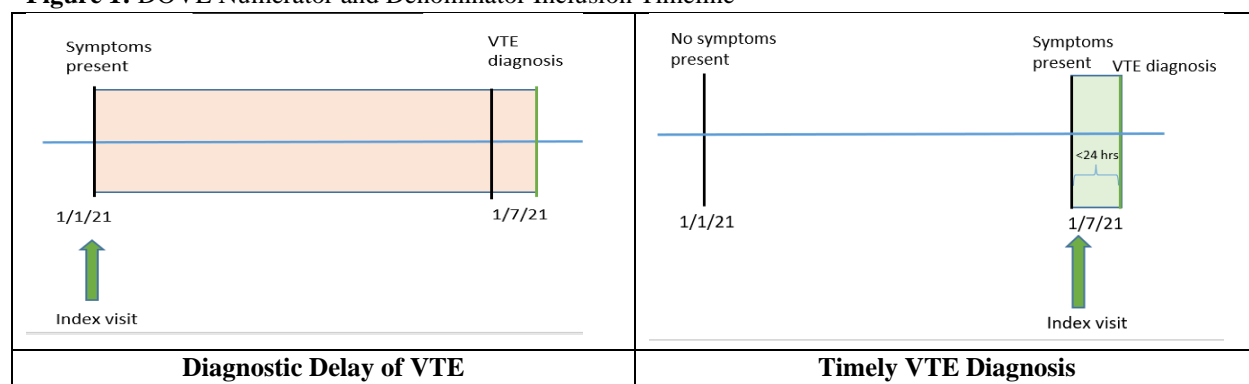| DOVE Inclusion Criteria	DOVE Exclusion Criteria
• Aged 18 years or older • PCP visit where VTE s/s are recorded • Diagnosis of VTE within 30 days of the primary care visit • No eligible VTE events within 6 months of the qualifying VTE event	• Receiving hospice within 90 days of the VTE encounter • Receiving palliative care within 90 days of the VTE encounter

Table 1: DOVE eCQM inclusion and exclusion criteria

occurred greater than 24 hours following the initial primary care visit within 30 days before diagnosis (**Figure 1**). Lower DOVE rates are indicative of better-quality performance. Inclusion and exclusion criteria are provided in **Table 1**.

Figure 1: DOVE Numerator and Denominator Inclusion Timeline

Diagnostic Delay of VTE	Timely VTE Diagnosis

Alpha Testing

Determining Data Sources and the Quality and Feasibility of Data Elements: Based on the measure specifications, we determined the specific data elements needed to calculate the measure. The DOVE eCQM required structured

and unstructured EHR data components. Structured data were needed to assess patient demographic characteristics, inclusion, and exclusion criteria, to confirm the VTE diagnoses, and to measure the time duration between the initial (index) primary care encounter and the VTE diagnosis date (**Figure 1**). Because patient s/s are commonly recorded in clinical notes, unstructured data from primary care notes were needed to identify VTE-related s/s recorded during the index primary care visit. Before finalizing the data sources for the DOVE eCQM, we used an iterative process to validate the data elements with our TEP and other stakeholders to ensure that they accurately reflected the DOVE concept being measured. Included in this process was developing a VTE phenotyping algorithm using the structured data to accurately identify patients with a true VTE diagnosis and an NLP algorithm to identify patients who presented to primary care with VTE s/s. This manuscript describes the DOVE eCQM measure testing. The details of the development and testing of the DOVE phenotyping (14) and NLP(17) algorithms are described elsewhere. Our NLP algorithm extracts a total of 28 VTE s/s identified by literature and expert consensus. The final VTE symptom list is available in **Table 2**.

Table 2: VTE-related S/s extracted using the NLP algorithm

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

Alpha testing involved analysis of data from both sites. To assess the feasibility of data extraction, we used the National Quality Forum (NQF) eCQM Feasibility Scorecard to assess the availability of the data elements needed to calculate the measure along the scorecard's four domains of data availability, accuracy, mapping to data standards, and required data collected during routine care (Table 3).

Table 3: Structured Data Elements needed to calculate Diagnostic Delay of VTE

• Age at VTE diagnosis • Sex	• Race/ethnicity • Insurance type	• Condition • Primary care encounter	• CPT imaging codes for VTE • VTE s/s at primary care encounter	• RxNorm therapeutic anticoagulant orders • ICD billing codes related to VTE

We also secured feedback and confirmed measure feasibility from our TEP, healthcare providers and other stakeholders through structured meetings, interviews, focus groups, and public comment. One area where feasibility was a concern was that the NLP algorithm returned "0" (no s/s) or "1" (s/s present). In order to use the CMS Measure Authoring Tool(19) which is required for measure development and submission, and for the eCQM to be classified as a fully specified measure, the NLP output needed to be converted to a national terminology standard that could be used to develop a value set (see **Figure 2** for how we addressed this challenge).

Face validity. We conducted face validity testing to demonstrate that the DOVE eCQM would be meaningful and beneficial to providers, patients, and informatics professionals, from the perspectives of stakeholders and experts in the field. We provided the TEP with several opportunities during the measure development process to suggest improvements and

Figure 2: Method for converting NLP s/s output into a standard value set. We mapped the VTE s/s identified by the NLP algorithm (**Table 2**), to SNOMED CT codes. Specifically, when the NLP algorithm identified the presence of a specific symptom, it resulted in a positive value ("1") populating a structured query language (SQL) query output table. The relevant SNOMED CT symptom code mapped to a new output table for the DOVE eCQM to access. The eCQM algorithm accesses the SNOMED CT code and uses that code to populate the measure numerator and denominator in the study population. The resulting VTE symptom value set was published in the Value Set Authority Center (VSAC)(1) and is available for all who implement the DOVE measure.

refinements to measure specifications to ensure optimal performance. The TEP provided feedback including timeframe for delayed diagnosis, approach for identifying incident cases of VTE, washout period to exclude chronic cases, measure numerator and denominator, exclusions, and performance of VTE phenotyping and NLP algorithms. Then, a formal face validity vote was conducted with our TEP via an online survey that was sent to TEP members by email after a DOVE eCQM 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: "Can the VTE Diagnostic Delay in Primary Care eCQM, as specified, be used to distinguish good from poor clinician group-level quality related to patient safety?" TEP members were blinded to each other's responses but were told the final result after all eligible members had voted.

Beta Testing

Reliability. Beta testing was conducted to determine an overall site-specific DOVE rate (sites 1 and 2),

and by provider groups within sites (site 1). As a non-interoperable and semi-rural site, site 2 technical experts faced difficulties in accurately capturing clinician group levels. Therefore, this site was assessed as a single clinician group at the facility level. Site 1 data were randomly split into two samples, a test sample and a validation sample. The demographics of the two samples were compared (using p values). The overall DOVE rate was calculated (sites 1 and 2) and in a sub-analysis that included provider groups with an average of 40 cases that met the denominator criteria (site 1). For site 1, we used a spearman correlation coefficient to compare the relative rankings of clinician groups in the test and validation samples.

Measure Validity Testing. Measure validity testing was completed by conducting multiple rounds of chart review by trained research staff on random samples of patients. Chart reviewers determined whether a patient met the criteria for inclusion in the numerator and/or denominator or whether they should be excluded. The manual chart review classifications (e.g., "gold standard") were compared to those produced by the eCQM. All discrepancies were analyzed and resolved through meetings with the research and development teams and the eCQM code was refined to address errors. A Kappa value was calculated after each round to quantify the level of agreement.

DOVE Screening Algorithm Validation. We further validated the DOVE eCQM algorithm by assessing the number of patient encounters in site 1 that were and were not captured within the measure across four domains:

1. Adult patients who had VTE related s/s during the primary care visit who received a diagnosis of VTE within 30 days of the primary care visit.
2. Adult patients who had VTE related s/s during the primary care visit who did NOT receive a diagnosis of VTE within 30 days of their primary care visit.
3. Adult patients who had NO VTE related s/s during the primary care visit who received a diagnosis of VTE within 30 days of the primary care visit.
4. Adult patients who had NO VTE related s/s during the primary care visit who did NOT receive a diagnosis of VTE within 30 days of the primary care visit.

For patients who received a diagnosis of VTE within 30 days of their primary care visit (with and without VTE s/s in primary care), we used the total sample of VTE encounters in the site 1 EHR enterprise database that met our criteria for a VTE diagnosis (e.g., the co-occurrence of ICD-10, RxNorm, and CPT codes related to VTE) from 2016-2021 (n=187,409). For patients with VTE s/s in primary care who did not receive a VTE diagnosis within 30 days, we used a random sample of 10,000 patients from the site 1 EHR enterprise system from 2016-2021 who did not receive a VTE diagnosis. The rationale for using a subsample is that our data extraction query identifies patients with a VTE diagnosis first, then moves backwards to assess VTE s/s during primary care via the NLP algorithm. As site 1 is a major U.S. healthcare system, assessing all patient notes for all primary care encounters during the 2016-2021 study cohort for VTE s/s was not feasible. We did not calculate the number of patients who did not have VTE s/s in primary care who did not receive a VTE diagnosis within 30 days as this is equivalent to the general population.

Rate calculation. We calculated the DOVE eCQM rate overall (facility level) in site 1. Due to limitations in group-level analysis, the DOVE rate was assessed at the facility level only in site 2. Using site 1 data, we also did a sub-analysis using data from the 15 practices that had an average of at least 40 encounters that met denominator inclusion criteria. The purpose of the sub-analysis was to assess how the eCQM performed in a variety of primary care facilities with enough cases to be both clinically and statistically meaningful. Also, in site 1, we conducted a sub-analysis to assess the DOVE performance rate after stratifying clinician groups into four cohorts by clinician group size. 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 without excluding practices that have relatively few VTE case encounters. Clinician groups were stratified into the following categories: >100 encounters during the study period, 50-99 encounters, 25-49 encounters, and <25 encounters.

Risk adjustment. Our team chose not to risk adjustment the DOVE eCQM rates because performing risk adjustment would establish a lower standard of care for individuals with risk adjusted characteristics which are unrelated to delayed diagnosis. The DOVE eCQM assesses the rate of encounters where a VTE is diagnosed >24 hours following a primary care visit where VTE-related s/s were documented in the EHR clinical notes. This means that at the time of the primary care visit, the clinician had the information necessary to investigate VTE based on self-reported patient s/s but did not take action in a timely manner, which is dangerous for the treatment and management of VTE. Additionally, VTE are relatively rare and dangerous events. Risk adjustment could impose sample size minimums at the clinician-group level which could 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. In conversations with our 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.

Establishing a benchmark. Because there are no existing measures or benchmarks for delayed diagnosis of VTE, the ABC method(20) was used to calculate a DOVE benchmark rate. The ABC method is suitable for comparing samples between groups of different sizes and establishes benchmark performance as the level consistently attained by the top performers accounting for at least 10% of the overall population. First, we calculated the adjusted performance for all groups and then ranked groups based on performance. We then calculated the top performing groups that accounted for at least 10% of the overall population (benchmark group). Finally, we calculated the ABC benchmark using the sum of all numerators in the benchmark group and the sum of all denominators in the benchmark group. This same process was used to calculate the benchmark rate in test and validation samples.

Results
Alpha Testing

The site 1 sample included a total of 214 primary care sites. The clinician groups in the site 1 sample with an average of 40 encounters that met denominator inclusion criteria had a total of 3,106 primary care encounters (range=11-1,046) that met the measure inclusion criteria. The site 2 sample represented a total of 245 encounters that met the measure inclusion criteria. Based on the NQF feasibility scorecard, 100% of the data elements needed to calculate the DOVE eCQM were available, accurate, coded using nationally accepted terminology standards, and captured by clinicians during the course of care without additional documentation burden in both sites. **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.

Table 4: Descriptive Statistics of the Included Sample (sites 1-2) and Reliability of Split Sample Demographics (Site 1)

	Site 1 (Overall)	Site 1 Test Sample	Site 1 Validation Sample	P Value	Site 2 (Overall)
Number of encounters, n	5514	1168	1177	NA	632
Encounters included in the measure, n	3591	1168	1177	NA	245
Encounters excluded from the measure, n	1923	0	0	NA	387
Number of delayed VTE diagnosis events, n	2607	847	883	NA	189
Site delayed diagnosis rate, %	72.60	72.52	75.02	NA	77.14
Number of clinician groups, n	214	15	15	NA	1
Age:					
Mean age at VTE (SD)	65.89 (15.27)	66.49 (14.58)	65.39 (15.05)	0.16	58.14 (N/A)
Age ≥65 (%)	2082 (57.98)	96 (59.59)	659 (55.99)		84 (34.29)
Age <65 (%)	1509 (42.02)	472 (40.41)	518 (44.01)		161 (65.71)
Sex (%):					
Female	1847 (51.43)	608 (52.05)	620 (52.68)	0.73	129 (52.65)
Male	1744 (48.57)	560 (47.95)	557 (47.32)		116 (47.35)
Self-Reported Race (%):					
Black/African American	312 (8.69)	131 (11.22)	110 (9.35)	0.95	24 (9.80)
White	2945 (82.01)	936 (80.14)	950 (80.71)		221 (90.20)
Other*	334 (9.30)	101 (8.65)	117 (9.94)		0 (0%)
Self-Reported Ethnicity (%):					
Hispanic	233 (6.49)	62 (5.31)	77 (6.54)	0.47	4 (1.63)
Non-Hispanic	3290 (91.62)	1,076 (92.12)	1,076 (91.42)		236 (96.33)
Missing/Declined	68 (1.89)	30 (2.57)	24 (2.03)		5 (2.04)
Insurance Type (%):					
Public Insurance	1969 (54.83)	645 (55.22)	638 (54.20)	0.94	162 (66.12)
Private Insurance	1611 (44.86)	523 (44.44)	539 (45.63)		83 (33.88)

Other insurance**	11 (0.31)	4 (0.34)	2 (0.17)		0 (0)
English as a first language (%)	3325 (92.59)	1,096 (93.84)	1,075 (91.33)	0.24	241 (98.37)
Median income (via ZIP Code), USD (SD)	74,359 (27,059)	73,800 (27,360)	73,610 (27,620)	0.46	38,254 (N/A)
Mean number of VTE s/s (SD)	2.31 (1.34)				1.4 (N/A)

*Other racial 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 encounters	5514	632
Encounters excluded from the measure, n	1923	387
Number of clinician groups	170	1
Age:		
Mean age at VTE (SD)	63.86 (15.26)	55.61 (N/A)
Age ≥65 (%)	1026 (53.35)	131 (33.85)
Age <65 (%)	897 (46.65)	256 (66.15)
Sex (%):		
Female	930 (48.36)	199 (51.42)
Male	993 (51.64)	188 (48.59)
Self-Reported Race (%):		
Black/African American	177 (9.20)	58 (14.99)
White	1571 (81.70)	326 (84.24)
Other*	175 (9.10)	3 (0.78)
Self-Reported Ethnicity (%):		
Hispanic	140 (7.28)	9 (2.33)
Non-Hispanic	1746 (90.80)	373 (96.38)
Missing/Declined	37 (1.92)	5 (1.29)
Insurance Type (%):		
Public Insurance	1009 (52.47)	298 (77.0)
Private Insurance	911 (47.37)	85 (21.96)
Other**	3 (0.16)	4 (1.03)
English as a first language (%)	1757 (91.37)	375 (96.90)
Median income (via ZIP Code), USD (SD)	73,823 (27,112)	38,965 (N/A)

*Other racial category includes Asian, American Indian or Alaska Native, and race self-reported as "other"

**Other insurance category includes free care and self-pay

Data element availability. Patient demographics were consistently present in the EHR (Table 6). Missing data was minimal, the only variable with missing data was patient ethnicity (less than 2%).

Table 6: Frequency of Data Elements (Denominator Sample)

	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				

Face validity. Five out of six (83%) TEP members were present for the face validity vote. The final vote was 5/5 (100%) agreement that the DOVE eCQM, as specified (**Figure 3**), can be used to distinguish good from poor clinician group-level quality related to patient safety.

Beta Testing

Reliability. We found no significant differences in patient sociodemographic characteristics (Table 3) in our randomly split test and validation samples for site 1 (*p*=0.16-0.95). DOVE rates were also similar across test and validation samples (see **Table 7**). 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 (r=.782, CI 0.43724, 0.94290). 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 calculated ICC=3.2% is above the NQF required level of >0.50%.

Figure 3: DOVE eCQM Measure Description *(Integrated care delivery systems, specified at the provider group level)*
Denominator: All adult patients (18+) diagnosed with VTE who presented to primary care with VTE-related s/s within the 30 days before VTE diagnosis.
Numerator: Subset of the denominator where the VTE diagnosis occurred greater than 24 hours following the initial primary care visit within 30 days before diagnosis.
Exclusions: Receiving hospice or palliative care within 90 days of the VTE encounter.

Measure Validity Testing. There was excellent agreement between the gold standard (manual chart reviewers) and the eCQM output. Following the manual chart review of 30 patients from site 1, 13 patients were sorted into the denominator, 9 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.

DOVE Screening Algorithm Validation. The results of our screening algorithm testing found a total of 6,620 primary care patient encounters where patient s/s were present, and the patient received a positive VTE diagnosis within 30 days. There were a total of 1,043 primary care patient encounters where no VTE s/s were reported, but these patients received a positive VTE diagnosis within 30 days. In the random sample of 10,000 primary care patient encounters of patients who did not receive a VTE diagnosis, 5,158 patients had VTE-related s/s present (recorded in primary care note). See **Table 7**.

Table 7: DOVE Screening Algorithm Validation

DOVE Denominator		
Aged 18 years or older on the date of the primary care visit. (Had primary care visit and is aged 18+)	DID receive a diagnosis of VTE within 30 days of their primary care visit	DID NOT receive a diagnosis of VTE within 30 days of their primary care visit
Have >1 VTE-related s/s	6,620/187,409 = 3.53%	5,158/10,000 = 51.6%
No VTE-related s/s documented	1,043/187,409 = 0.56%	N/A (general population)

DOVE Rate Calculation. The overall DOVE rate in site 1 across all 214 primary care sites was 72.60% (SD=0.22). Across the largest 15 primary care practices in site 1 where an average of 40 patients met the denominator criteria, the DOVE rate was 73.74% (SD=0.09) with group-level rates ranging from 50.00%-85%. The DOVE rate in the test and validation samples ranged from 50.00%-84.21% and 57.14%- 85.00% respectively. The overall DOVE rate in site 2 was 77.14% (see Table 8).

Table 8: Rate Calculation for Sites Included in the Test and Validation Samples (Site 1)

Primary Care Site	**Test Sample**			**Validation Sample**		
	Denominator Encounters	Numerator Encounters	Site Rate	Denominator Encounters	Numerator Encounters	Site Rate
Practice A	19	16	84.21%	20	17	85.00%
Practice B	20	12	60.00%	21	12	57.14%
Practice C	20	14	70.00%	21	16	76.19%
Practice D	21	16	76.19%	22	17	77.27%
Practice E	21	13	61.90%	22	18	81.82%
Practice F	22	15	68.18%	22	17	77.27%
Practice G	29	23	79.31%	29	23	79.31%

Practice H	30	23	76.67%	30	25	83.33%
Practice I	32	19	59.38%	32	21	65.63%
Practice J	34	23	67.65%	35	26	74.29%
Practice K	44	22	50.00%	44	26	59.09%
Practice L	56	41	73.21%	57	46	80.70%
Practice M	59	47	79.66%	60	50	83.33%
Practice N	237	180	75.95%	239	197	82.43%
Practice O	523	383	73.23%	523	372	71.13%
Total	**1,168**	**847**	**72.52%**	**1,177**	**883**	**75.02%**

In the tiered subgroup analysis of site 1, rates between cohorts ranged from 65.78%-74.99%, and variation within cohorts is seen in each cohort rate range (**Table 9**).

Table 9: Sub Analysis by Clinician Group Sample Size

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

DOVE Benchmark. Based on the ABC method(20) which establishes benchmark performance as the level consistently attained by the top performers accounting for at least 10% of the overall population, the overall benchmark rate for the DOVE eCQM is 49.63%. The benchmark rate for the split (test and validation) samples was 69.97% and 69.52% respectively.

Discussion and Conclusions

In this study, we conducted alpha and beta testing of the DOVE eCQM in two geographically distinct healthcare systems using different EHR vendor systems. We found that the DOVE rate was high (>70%) in both systems. We found that measure implementation was feasible; data elements needed to calculate the measure are routinely available within the EHR. Based on our testing, the DOVE eCQM is most feasible for use in integrated delivery networks where patients seen in primary care receive their follow-up care in that same system. When tested in a large non-integrated care delivery system, many patients received their follow-up care out of network where EHR systems were not interoperable. The measure was less feasible in this setting, and we determined that the measure may be most useful for quality measurement when implemented in an integrated care delivery network.

Currently there are no federal level measurement tools in place to track VTE events, or delayed diagnosis of VTE so the DOVE rates identified in two geographically different U.S. healthcare systems cannot be compared against an existing metric. To assess clinically and practically meaningful differences in performance measure scores among our samples, we conducted a sub-analyses of clinician groups with an average of 40 cases in the denominator and by stratifying all clinician groups by encounter sample sizes into four cohorts and assessed overall DOVE rates and ranges. We found high DOVE rates with some variation in each cohort. This demonstrates that the DOVE eCQM may be clinically and practically meaningful for understanding delayed diagnosis rates across different 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 and facility levels. We also calculated a DOVE benchmark based on best performing sites in our sample that can be used by other systems who implement the measure. In addition to providing benefit within a payment program, this measure could serve as the first passive monitoring system to assess delayed diagnosis of VTE at the national level.

Our team chose not to risk adjustment the DOVE eCQM rates. In the 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(21, 22). Risk of VTE is associated with older age (23, 24). African American race is associated with higher rates of VTE complications compared to white race(25). Although there are some disparities in the individuals who experience VTE, there should not be social disparities in the delayed diagnosis of VTE following the onset of s/s noted by a physician. The goal of this measure is to quantify

and reduce delayed VTE events. Risk adjustment could potentially mask the rate of delayed events among vulnerable populations. In the future, we can use the model predictors to calculate expected rates for clinician groups who use the eCQM to compare against the observed rate.

Finally, the goal of the DOVE eCQM is not to increase imaging to rule out VTE for patients who present to primary care with s/s but rather to identify practices that could benefit from education on current VTE diagnosis guidelines. The guideline on *Diagnosis of Venous Thromboembolism(26)* developed by the American Society of Hematology (2018, reviewed 2022) and endorsed by the American Academy of Family Physicians recommends the use of a D-dimer test to rule out DVT followed by imaging for patients requiring additional testing. To achieve our objective of reducing delays in VTE diagnosis and enhancing clinical practice, our team is in the process of developing clinical decision support (CDS) that uses NLP and machine learning methods to identify patients with a high likelihood of current VTE during a primary care visit. Together, the CDS and the DOVE eCQM could improve the overall quality of clinical practice and reduce delays in VTE diagnosis without causing unnecessary and expensive imaging.

Limitations. The site 2 sample size was too small to stratify by clinical group and the site 2 technical experts faced difficulties in accurately capturing distinct clinician groups within this semi-rural healthcare system. Therefore, clinician group levels could not be calculated, and this site was assessed as a single clinician (provider) group at the facility-level. This is not a limitation of the DOVE eCQM, as it can be overcome by using clinician group Tax Identification Numbers (TIN) to identify primary care provider groups, if the measure were implemented nationally (required for all eCQMs used in the National Payment Program). We did not have access to TIN numbers for this project. A second issue which is a limitation of eCQMs in general is that complete and accurate measurement requires interoperable EHRs systems. This measure includes an index primary care visit and a VTE event within 30 days. Patients who receive their primary care visit in one system but travel out of network for VTE diagnosis and/or treatment (e.g., a patient who is asked by their primary care provider to go the emergency room (ER) for an imaging scan to rule out VTE and the patient goes to the ER closest to their home which is out of network) would not be included in the measure since the in network EHR would include the ICD 10 codes but not the CPT or RX norm codes that are captured in the facility doing the imaging and treatment.

As noted in **Table 6**, the DOVE eCQM excludes a large amount of diagnosed VTEs. The majority of VTEs were excluded because they did not occur within 30 days of a primary care encounter. Patients with s/s during a primary care visit were greater than six times more likely to have their VTE diagnosed within 30 days than those who presented without s/s. Over half of the sample of patients without VTE presented in primary care with VTE-related s/s. This is expected because the s/s of VTE are nonspecific and very similar to the s/s of many acute and chronic illnesses routinely managed and treated by primary care providers. VTE can be difficult to diagnose, and the goal of this eCQM is to identify cases where a primary care provider had information available to identify a VTE and did not within a defined time window. Non-symptomatic VTE are not included in this measure as there is no information available to the primary care provider to signal a potential VTE. Our second testing site was not an integrated care network and did not have an interoperable EHR across sites, thus many patients were excluded from the measure. Although this measure is intended for use in integrated delivery networks, the high DOVE rate in site 2 suggests that this measure can also be used for quality improvement purposes in non-integrated healthcare systems. Additional work is needed to establish benchmarks in these types of systems. Finally, missing data is minimal (ethnicity only and <2%) and is not expected to bias the results of the measure because it is not risk adjusted or stratified by ethnicity.

The rate of delayed VTE diagnosis in patients seen within 30 day of diagnosis in primary care is unacceptably high. The DOVE eCQM provides a way to quantify VTE diagnostic delay and can be used for both quality reporting (integrated care delivery systems) and quality improvement purposes. Our team is continuing testing with a third external site and submitting the DOVE eCQM for consideration in the National Quality Payment program for use in integrated care delivery systems. 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.

References
1. NIH. Value Set Authority Center 2023 [Available from: https://vsac.nlm.nih.gov/.
2. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med. 2009;169(20):1881-7.

3. 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;126(9):832 e13-21.
4. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous thromboembolism: a public health concern. Am J Prev Med. 2010;38(4 Suppl):S495-501.
5. Nijkeuter M, Sohne M, Tick LW, Kamphuisen PW, Kramer MH, Laterveer L, et al. The natural course of hemodynamically stable pulmonary embolism: Clinical outcome and risk factors in a large prospective cohort study. Chest. 2007;131(2):517-23.
6. Perrier A, Roy PM, Aujesky D, Chagnon I, Howarth N, Gourdier AL, et al. Diagnosing pulmonary embolism in outpatients with clinical assessment, D-dimer measurement, venous ultrasound, and helical computed tomography: a multicenter management study. Am J Med. 2004;116(5):291-9.
7. van Strijen MJ, de Monye W, Schiereck J, Kieft GJ, Prins MH, Huisman MV, et al. Single-detector helical computed tomography as the primary diagnostic test in suspected pulmonary embolism: a multicenter clinical management study of 510 patients. Ann Intern Med. 2003;138(4):307-14.
8. Karwinski B, Svendsen E. Comparison of clinical and postmortem diagnosis of pulmonary embolism. J Clin Pathol. 1989;42(2):135-9.
9. Uhland H, Goldberg LM. Pulmonary Embolism: A Commonly Missed Clinical Entity. Dis Chest. 1964;45:533-6.
10. Dalen JE. Pulmonary embolism: what have we learned since Virchow?: treatment and prevention Chest 2002;122.5 1801-17.
11. 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.
12. CMS.gov. Electronic Clinical Quality Measures Basics 2023 [February 18, 2023]. Available from: https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/clinicalqualitymeasures.
13. CMS. CMS Measures Inventory Tool (2023) Search: "eCQM" 2023 [Available from: https://cmit.cms.gov/CMIT_public/ListMeasures?struts.token.name=token&token=2BBIEBF5W0JF7G0DM1W9WQ6B U3AH2TEJ&filters=Status_untok%7Cd%7CActive&view=&makeStick=&wasSearchSubmitted=true&q=ecqm
14. Pullman A, Dykes P, Syrowatka A, Coghlan E, Song W, Sainlaire M, editors. Defining Venous Thromboembolism Using the Electronic Health Record: A Data Driven Approach American Medical Informatics Association Annual Summit; 2022; Chicago, IL.
15. Fang MC, Fan D, Sung SH, Witt DM, Schmelzer JR, Steinhubl SR, et al. Validity of Using Inpatient and Outpatient Administrative Codes to Identify Acute Venous Thromboembolism: The CVRN VTE Study. Med Care. 2017;55(12):e137-e43.
16. 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;20(4):364-8.
17. Laurentiev J, Bowen M, Pullman A, Song W, Syrowatka A, Chen J, et al. Development and Optimization of an Extraction Tool for Identifying Signs and Symptoms of Venous Thromboembolism in Primary Care Clinical Notes. Under review. 2023.
18. Laurentiev J, Pullman A, Song W, Syrowatka A, Sainlaire M, Chang Y, et al., editors. Development and Validation of an Extraction Tool for Identifying Signs and Symptoms of Venous Thromboembolism in Primary Care Clinical Notes. American Medical Informatics Association Annual Symposium; 2022; Washington, DC.
19. NQF. NQF Measure Authoring Tool 2023 [Available from: https://www.qualityforum.org/mat/.
20. Kiefe CI, Weissman NW, Allison JJ, Farmer R, Weaver M, Williams OD. Identifying achievable benchmarks of care: concepts and methodology. Int J Qual Health Care. 1998;10(5):443-7.
21. Mansour S, Alotaibi G, Wu C, Alsaleh K, McMurtry MS. Sex disparities in hospitalization and mortality rates for venous thromboembolism. J Thromb Thrombolysis. 2017;44(2):197-202.
22. Marshall AL, Bartley AC, Ashrani AA, Pruthi RK, Durani U, Gonsalves WI, et al. Sex-based disparities in venous thromboembolism outcomes: A National Inpatient Sample (NIS)-based analysis. Vasc Med. 2017;22(2):121-7.
23. Anderson FA, Jr., Wheeler HB, Goldberg RJ, Hosmer DW, Patwardhan NA, Jovanovic B, 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;151(5):933-8.
24. Gillum RF. Pulmonary embolism and thrombophlebitis in the United States, 1970-1985. Am Heart J. 1987;114(5):1262-4.
25. Aujesky D, Long JA, Fine MJ, Ibrahim SA. African American race was associated with an increased risk of complications following venous thromboembolism. J Clin Epidemiol. 2007;60(4):410-6.
26. Lim W, Le Gal G, Bates SM, Righini M, Haramati LB, Lang E, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: diagnosis of venous thromboembolism. Blood Adv. 2018;2(22):3226-56.