# 2024 MIPS Peer-Reviewed Journal Article Requirement Template

Section 101(c)(1) of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) requires submission of new measures for publication in applicable specialty-appropriate, peer-reviewed journals prior to implementing in the Merit-based Incentive Payment System (MIPS). Such measures will be submitted by the Centers for Medicare & Medicaid Services (CMS), to a journal(s), before including any new measure on the MIPS Quality Measures List. The measure submitter shall provide the required information for article submission under the MACRA per the MIPS Annual Call for Quality Measures submission process.

Interested parties submitting measures for consideration through the MIPS Annual Call for Quality Measures must complete the required information by the CMS Annual Call for Measures deadline (8 p.m. ET on May 10, 2024). Some of the information requested below may be listed in specific fields in the CMS Measures Under Consideration (MUC) Entry/Review Information Tool (MERIT); however, to ensure that CMS has all of the necessary information and avoid delays in the evaluation of your submission, please fully complete this form as an attached Word document. The information in MERIT must be consistent with the information below, including the following, but not limited to:

* **Title:** Diagnostic Delay of Venous Thromboembolism (DOVE) electronic clinical quality measure (eCQM)
* **Meaningful Measures 2.0 Framework Domain:** Patient Safety

**Measure Steward:** Brigham and Women’s Hospital

**Measure Developer:** Brigham and Women’s Hospital

**Description:** Percentage of adult patients with a venous thromboembolism who’s VTE was diagnosed >24 hours following a primary care encounter where symptoms were present (within 30 days).

1. **Statement**

* **Background (Why is this measure important?).**

Venous thromboembolism (VTE) includes pulmonary embolism and deep vein thrombosis. VTE is a commonly missed or delayed diagnosis (Schiff et al., 2009; Tagalakis et al., 2013) affecting approximately 300,000-600,000 individuals in the U.S. each year, and requiring timely and adequate treatment to decrease mortality (Beckman et al., 2010). The 30-day mortality rate is 23% (Tagalakis et al., 2013; Nijkeuter et al., 2007; Perrier et al., 2004; van Strijen et al., 2003). 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 (Karwinksi, et al. 1989, Carvalho Bricola, et al. 2013). Earlier diagnoses of VTE may reduce the morbidity and mortality associated with this dangerous condition (Bhatt et al., 2020) and promote patient safety.

While patients often report symptoms to their primary care provider, VTE is also a commonly missed or delayed diagnosis in primary care settings (Walen et al., 2016). 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 (Walen et al., 2016). 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 (CMS, 2023). 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. 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.

**References:**

1. 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.
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;126(9):832 e13-21.
3. Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous thromboembolism: a public health concern. Am J Prev Med. 2010;38(4 Suppl):S495-501.
4. 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.
5. 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.
6. 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.
7. Karwinski B, Svendsen E. Comparison of clinical and postmortem diagnosis of pulmonary embolism. J Clin Pathol. 1989;42(2):135-9.
8. 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.
9. 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>
10. 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.
11. CMS.gov. Electronic Clinical Quality Measures Basics 2023 [February 18, 2023]. Available from: <https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/clinicalqualitymeasures>.

* **Environmental scan (Are there existing measures in this area?).**

At this time there are no CMS-endorsed measures (eCQM or claims-based) related to the diagnostic delay of VTE. Considering the dangers of this missed diagnosis and the benefits of early detection, this eCQM can fill the existing gap in VTE-detection measures and promote patient safety.

1. **Gap Analysis**

* **Provide evidence for the measure (What are the gaps and opportunities to improve care?).**

|  |
| --- |
| Diagram, text  Description automatically generated  **Figure 1:** DOVE Logic Model |

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).

**References:**

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.

* **Expected outcome (patient care/patient health improvements, cost savings).**

The expected outcome for this eCQM is a reduction in delayed diagnosis of VTEs in the primary care setting by providing clinician groups with their quantified DOVE rates. The eCQM development team is also developing a Clinical Decision Support (CDS) tool to work alongside the eCQM; the goal of this CDS tool is to help primary care providers identify a VTE based on patient-reported symptoms in real time to assist in a DOVE rate reduction at the clinician group level. Reducing the incidence of delayed diagnosis of VTE can promote patient safety and outcomes as well as reduce healthcare costs associated with the increased morbidity and mortality of delayed diagnoses (Ruppert et al., 2011).

**References:**

1. 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.

* **Recommendation for the measure (Is it based on a study, consensus opinion, USPSTF recommendation etc.?).**

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). This gap in VTE detection has not been addressed at the federal level in the 15 years following this declaration.

It is well known that VTEs should be treated immediately upon identification, and that delays in treatment can be dangerous for patients (Liederman et al., 2020). Although there are many clinical guidelines and systematic reviews relating to VTEs that were reviewed during measure development (Xiang et al., 2019; Bhatt et al., 2020; Khan et al., 2021; Becattini et al., 2016; THANZ, 2019; Qaseen et al., 2007), these guidelines and reviews focus on treatment and identification approaches rather than reiterate the need for the timeliness of identification. Supported by the Gordon and Betty Moore Foundation (GBMF) under their Diagnostic Excellence Initiative, our team completed a series of studies to improve VTE diagnostic performance. The GBMF’s Diagnostic Excellence Initiative focuses on VTE and other vascular events because sub-optimal diagnosis is responsible for a disproportionate share of serious harm and preventable death.  Access to validated VTE diagnostic performance measures **will provide an opportunity to improve diagnostic performance** and to build the measurement infrastructure needed to systemically measure VTE diagnostic performance in real time, thus improving the ability to quantify performance and guide improvements.

**References:**

1. Liederman Z, Chan N, Bhagirath V. Current challenges in diagnosis of venous thromboembolism. Journal of Clinical Medicine. 2020 Oct 29;9(11):3509.
2. 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>
3. 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>
4. 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>
5. 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/>
6. New guidelines from the Thrombosis and Haemostasis Society of Australia and New Zealand for the diagnosis and management of venous thromboembolism, 2019.Available at: <https://www.thanz.org.au/documents/item/414>
7. 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/>
8. **Reliability/Validity**

* **What testing has been performed at the level of implementation? (MIPS requires full measure testing at the individual clinician level (and may also need to be tested at the group level) for MIPS Clinical Quality Measures (CQMs) and Electronic Clinical Quality Measures (eCQMs) collection types. Administrative claims measures tested at the group level require a reliability threshold to be implemented at the group level.) Please provide testing results including the N value, Bonnie test case results, correlation coefficient and any other pertinent information or values to be considered.**

**Measure Numerator:** 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).

**Measure Denominator:** All adult patients (age 18 years and older) presenting in primary care with VTE-related symptoms (see **Table 1**) with an eligible lower limb VTE event (see below) are included in the measure denominator. 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 clinical notes by a natural language processing (NLP) algorithm.

Criteria for an eligible VTE event:

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 a lower limb Venous Thromboembolism within 30 days of their primary care visit. For a patient to have a VTE diagnosis, they must have all of the following VTE-related codes within the same encounter:

* ICD-10 CM code for VTE.
* CPT codes for an imaging scan for VTE linked to the same encounter as the ICD-10 CM code.
* RxNorm order for therapeutic anticoagulants placed in the same encounter as the imaging scan.

4. Have no eligible VTE events within 6 months of the qualifying VTE event

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

*\*Intentionally Blank*

**Exclusions:** This measure excludes patients who have a hospice or palliative care encounter within 90 days of the eligible VTE encounter. 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. Value sets for denominator exclusions can be found in **Table 2**.

**Table 2: Value Sets for Measure Exclusion Criteria**

|  |  |  |
| --- | --- | --- |
| **Value Set Name** | **Steward** | **OID Number** |
| Hospice care | Brigham and Women’s Hospital | *2.16.840.1.113762.1.4.1108.15* |
| Palliative care | Brigham and Women’s Hospital | *2.16.840.1.113883.3.464.1003.101.12.1090* |

**Risk adjustment:** this measure is not risk adjusted.

**Stratification:** this measure is not stratified.

**Bonnie test case results:** 21/21 Bonnie testing patients passed with 100% coverage.

The Site 1 sample included a total of 214 primary care sites and Site 3 included 19. 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.

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

**Table 3:** Descriptive Statistics of the Included Sample

**Table 1: Included Sample**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Site 1** | **Site 2** | **Site 3** |
| Number of encounters | 5514 | 632 | 545 |
| Encounters included in the measure | 3591 | 245 | 292 |
| Encounters excluded from the measure | 1923 | 387 | 253 |
| Number of delayed VTE diagnosis events | 2607 | 189 | 239 |
| Site delayed diagnosis rate | 72.60% | 77.14% | 81.85% |
| Number of clinician groups | 214 | 1 | 19 |
| **Age:** |  |  |  |
| Mean age at VTE (SD) | 65.89 (15.27) | 58.14 (N/A) | 65.72 (16.42) |
| Age >65 (%) | 2082 (57.98%) | 84 (34.29%) | 163(55.82%) |
| Age <65 (%) | 1509 (42.02%) | 161 (65.71%) | 129(44.18%) |
| **Sex (%):** |  |  |  |
| Female | 1847 (51.43%) | 129 (52.65%) | 150(51.37%) |
| Male | 1744 (48.57%) | 116 (47.35%) | 142(48.63%) |
| **Self-Reported Race (%):** |  |  |  |
| Black/African American | 312 (8.69%) | 24 (9.80%) | 22(7.53%) |
| White | 2945 (82.01%) | 221 (90.20%) | 248(84.93%) |
| Other\* | 334 (9.30%) | 0 (0%) | 22(7.53%) |
| **Self-Reported Ethnicity (%):** |  |  |  |
| Hispanic | 233 (6.49%) | 4 (1.63%) | 8(2.274%) |
| Non-Hispanic  Multiple | 3290 (91.62%)  0 | 236 (96.33%)  0 | 281(96.23%)   3(1.03%) |
| Missing/Declined | 68 (1.89%) | 5 (2.04%) | 0 |
| **Insurance Type (%):** |  |  |  |
| Public Insurance | 1969 (54.83%) | 162 (66.12%) | 207(70.89%) |
| Private Insurance | 1611 (44.86%) | 83 (33.88%) | 80(27.4%) |
| Other\*\* | 11 (0.31%) | 0 (0%) | 5(1.71%) |
| English as a first language (%) | 3325 (92.59%) | 241 (98.37%) | 286(97.95%) |
| Median income (via ZIP Code) (SD) | $74,359 ($27,059) | $38,254 (N/A) | $58,348 ($9,997) |
| Mean number of VTE symptoms (SD) | 2.31 (1.34) | 1.4 (N/A) | 2.92 (1.78) |
| *\*Other racial category includes Asian, American Indian or Alaska Native, and race self-reported as "other"* | | | |
| \*Other insurance category includes free care from the hospital and self-pay | | |  |

**Table 4:** Descriptive Statistics of the Excluded Sample

**Table 2: Excluded Sample**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Site 1 (MGB)** | **Site 2 (UK)** | **Site 3 (PSH)** |
| Number of encounters | 5514 | 632 | 545 |
| Encounters excluded from the measure | 1923 | 387 | 253 |
| Number of clinician groups | 170 | 1 | 22 |
| **Age:** |  |  |  |
| Mean age at VTE (SD) | 63.86 (15.26) | 55.61 (N/A) | 62.66 (N/A) |
| Age >65 (%) | 1026 (53.35%) | 131 (33.85) | 132(52.17%) |
| Age <65 (%) | 897 (46.65%) | 256 (66.15) | 121(47.83%) |
| **Sex (%):** |  |  |  |
| Female | 930 (48.36%) | 199 (51.42) | 147(58.1%) |
| Male | 993 (51.64%) | 188 (48.59%) | 106(41.9%) |
| **Self-Reported Race (%):** |  |  |  |
| Black/African American | 177 (9.20%) | 58 (14.99) | 21(8.3%) |
| White | 1571 (81.70%) | 326 (84.24) | 203(80.24%) |
| Other\* | 175 (9.10%) | 3 (0.78) | 29(11.46%) |
| **Self-Reported Ethnicity (%):** |  |  |  |
| Hispanic | 140 (7.28%) | 9 (2.33) | 6(2.37%) |
| Non-Hispanic  Multiple | 1746 (90.80%)  0 | 373 (96.38)  0 | 243(96.05%)  4(1.58%) |
| Missing/Declined | 37 (1.92%) | 5 (1.29) | 0 |
| **Insurance Type (%):** |  |  |  |
| Public Insurance | 1009 (52.47%) | 298 (77.0) | 174(68.77%) |
| Private Insurance | 911 (47.37%) | 85 (21.96) | 77(30.43%) |
| Other\*\* | 3 (0.16%) | 4 (1.03) | 2(0.79%) |
| English as a first language (%) | 1757 (91.37%) | 375 (96.90) | 244(96.44%) |
| Median income (via ZIP Code) (SD) | $73,823 ($27,112) | $38,965 (N/A) | $58,890 ($8,926) |
| *\*Other category includes Asian, American Indian or Alaska Native, and race self-reported as "other"* | | |  |
| *\*Other insurance category includes free care from the hospital and self-pay (Site 1), or missing (Site 2)* | | | |

* + **Reliability Testing Results at the accountable entity level**

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 (clinician and group level analysis)

**NLP Phenotyping Algorithm Accuracy Testing:**

**Methods:** 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.

**Results:** 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*

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.

**Table 5:** 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 6:** 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) |

**Interpretation:** 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:**

**Methods:**

|  |
| --- |
| Graphical user interface, text, application, chat or text message  Description automatically generated  **Figure 2:** Novel VTE Phenotype cohort development and analytic pipeline |

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 (**Figure 2**):

* 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).

**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.

**Results:**

**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 7:** Value set codes used to indicate a VTE

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

|  |  |
| --- | --- |
| **Table 8:** Chart Review Results | |
| Data element | Accuracy |
| PPV | 95.80% |
| NPV | 100% |
| Sensitivity | 100% |
| Specificity | 95.69% |

**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.

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

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

* DOVE 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%

Graphical user interface

Description automatically generated with medium confidence

**Figure 3:** Alternative VTE Cohorts

**Signal to Noise Analysis:**

**Methods:** A signal to noise analysis was conducted for the practices at the individual clinician level (n=29) in for Site 1 and the clinician group level (n=43) in Site 1 and Site 3. 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.

**Results:** At the clinician group level, 43 groups from Site 1 and Site 3 were sampled and the median signal-to-noise statistical result was 0.3958 (95% CI; 0.3385, 0.4532). The minimum SNR was 0.2474 and the maximum was 0.9500. Analysis at the clinician level was conducted only for site 1, as it met the criterion of having a minimum of 10 patients per clinician. Sites 2-3 did not have enough clinicians with at least 10 patients in the denominator to complete the signal-to-noise analysis. At Site 1, 29 groups were sampled and the median signal-to-noise statistical result was 0.1763 (95% CI: 0.1544, 0.1983). The minimum SNR was 0.1225, the maximum was 0.3262. The SNR was stronger at the clinician group level compared to the clinician level, reinforcing our rationale as to why this measure is specified at the group level.

**Interpretation:** 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 43, 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.
   * **Face Validity Testing Results, Clinician Sites**

Throughout measure conceptualization, development, and refinement, we worked with a technical expert panel (TEP). The TEP consists of 6 members, three clinicians, one EHR expert, and two patient perspectives, all of whom were identified and recruited by the Site 1 primary investigator.

**DOVE TEP Members:**

* Jason Adelman (Columbia University)
* David W. Bates (Brigham and Women’s Hospital, clinician)
* Gregory Piazza (Brigham and Women’s Hospital, clinician)
* Isbelia Briceno (Oracle Cerner, EHR expert)
* Martie Carnie (Brigham and Women’s Hospital, patient perspective)
* Tania Powell (Brigham and Women’s Hospital, patient perspective)

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.

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.

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 in agreement with the voting statement among present members. 1 member was absent and did not vote. 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.

* + **Empiric Validity Testing Results at the accountable entity level**

**Methods:** 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.

**Results:** 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 9:** 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** |
| *\*Other category includes Asian, American Indian or Alaska Native, and race self-reported as "other"*  *\*\*Other insurance category includes free care and self-pay* | | | |

**Interpretation:** 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).

* + **Data Element/Patient Encounter Level Testing**

**Methods:** 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.

**Results:** 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.

**Interpretation:** 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.

* + **Exclusion Frequency**

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 10**).

**Table 10:** 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.

* + **What were the minimum sample sizes used for reliability results?**

There are no sample size minimums defined in the DOVE eCQM. The random split half correlation used a minimum sample size of 40 encounters (20 in the test sample, 20 in the validation sample).

* + **Other Information**
* **Is it risk adjusted? If so, how?**

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.

A sub analysis was performed to assess disparities by social determinant of health variables in Site 1 patients. A T-test assuming equal variances was performed to assess if there was significant variation between the subsamples (See **Table 11**). 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. 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 11:** Sub Analysis by Social Risk Factors (Site 1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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* | | | | |

* **What benchmarking information is available?**

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.

**Methods:** 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).

**Results:** 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%. 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.

**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.

* **Collection Type: Specify the data collection type.**

This eCQM leverages routinely collected electronic health record (EHR) data.

* **Specify measure stage of development.**

This measure is fully developed and testing in three U.S. healthcare systems has been completed.

* **For Patient Reported Outcome Performance Measures:**

N/A, this eCQM is not a PRO-PM.

1. **Endorsement**
   * **Provide the Consensus-Based Entity (CBE) (i.e., Partnership for Quality Measures (PQM)) endorsement status (and CBE ID) and/or other endorsing body. If the measure is only endorsed for paper records, please note endorsement for only the data source being submitted.**

The DOVE eCQM was submitted to PQM (Battelle) for CBE endorsement on 5/3/23. Our submission has been rejected.

1. **Summary**

* **Alignment with CMS Meaningful Measures Initiative or MACRA (if applicable).**

2017 Meaningful Measures Initiative aims to identify measures that minimize the level of burden for providers, provide significant opportunity for improvement, and address areas that safeguard public health (CMS, 2022), all of which are fulfilled by the DOVE eCQM. Per the MACRA Cascade of Measures tool (introduced in 2022), the DOVE eCQM can fit into the Diagnostic Accuracy/Error objective of the Reduction in National Serious Safety Events goal, as well as the Culture of Safety objective within the Safety Culture goal (CMS, 2022).

**References:**

1. Center for Medicare & Medicaid Services, 2022. Meaningful Measures Initiative [Internet]. Available at: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy>

* **Relevance to MIPS or other CMS programs.**

The DOVE eCQM is relevant to the MIPS program and other CMS QPPs as this eCQM is designed to provide a pay for performance aspect based on VTE identification into a provider group’s Medicare Part B reimbursement. The MIPS program was developed to reward eligible clinician groups for providing high quality care to patients, which in this measure refers to clinician groups who appropriately identify VTE’s based on patient symptoms and refer patients to VTE treatment.

* **Rationale: Use of measure for inclusion in program (specialty society, regional collaborative, other).**

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). 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 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.

**References:**

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. 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.
3. *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.
4. 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.
5. 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.

* **Public reporting (if applicable).**

N/A, this is a new measure. 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.

* **Preferable relevant peer-reviewed journal for publication.**

A manuscript on the testing of the DOVE eCQM has been submitted to the American Medical Informatics Association (AMIA) 2023 Annual Symposium.

* **Rationale as to how the measure correlates to existing cost measures and improvement activities, as applicable and feasible.**

The proposed eCQM correlates to three existing improvement activities, and two related cost measures.

**Improvement activities:**

The Merit-based Incentive Payment System (MIPS) incentivizes MIPS eligible clinicians for providing high quality care across quality, improvement, promoting interoperability and cost performance domains (REF). The proposed DOVE eCQM correlates well with a number of the 2023 MIPS high and medium weighted improvement activities from the Population Management, Care Coordination, and Patient Safety and Practice Assessment subcategories (CMS, 2023).

1. Population Management

* Anticoagulant Management Improvements (IA\_PM\_2): systematic anticoagulation management, including use of electronic decision support
* Care management for high-risk patients (IA\_PM\_14): longitudinal care management to patients at high risk for adverse health outcome or harm

1. Care Coordination

* Implementation of Use of Specialist Reports Back to Referring Clinician or Group to Close Referral Loop (IA\_CC\_1): providing specialist reports back to the referring individual MIPS eligible clinician or group to close the referral loop
* Implementation of improvements that contribute to more timely communication of test results (IA\_CC\_2): Timely communication of test results defined as timely identification of abnormal test results with timely follow-up.

1. Patient Safety and Practice Assessment

* Use of Patient Safety Tools (IA\_PSPA\_8): Use tools that assist specialty practices in tracking specific measures that are meaningful to their practice.
* Measurement and improvement at the practice and panel level (IA\_PSPA\_18): Measure and improve quality at the practice and panel level
* Implementation of formal quality improvement methods, practice changes, or other practice improvement processes (IA\_PSPA\_19): Adopt a formal model for quality improvement and create a culture in which all staff, including leadership, actively participates in improvement activities that could include designate regular team meetings to review data and plan improvement cycles

**Cost measures:** This measure is linked to the Medicare Spending per Beneficiary (clinician) cost measure. While the MSPB Clinician measure focuses on inpatient care, it would likely be correlated with the DOVE eCQM because delayed diagnosis of VTE is associated with increased morbidity and mortality (Dalen et al., 2002; Ozsu et al., 2011), and nearly 30% of untreated DVTs will result in severe swelling or ulceration of the leg (Raju et al., 1986; Benotti et al., 1984) which drives up hospitalization and increased healthcare spending (Ruppert et al., 2011). In addition, the Clinician and Total Per Capita Cost (TPCC) measure may also be correlated with the DOVE eCQM because delayed VTE diagnosis is associated with increased hospitalization and healthcare spending (Ruppert et al., 2011).

**References:**

1. Centers for Medicare & Medicaid Services. 2023 MIPS Improvement Activities User Guide. [Internet]. Available at: <https://qpp.cms.gov/mips/improvement-activities>
2. Centers for Medicare & Mediaid Services. Improvement Activities Performance Category: Traditional MIPS Requirements PY 2023. [Internet]. Available at: <https://qpp.cms.gov/content-management/node/2104>
3. 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.
4. Dalen JE. Pulmonary embolism: what have we learned since virchow?: natural history, pathophysiology, and diagnosis. Chest. 2002 Oct 1;122(4):1440-56.
5. 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.
6. Raju S, Fredericks RK. Late hemodynamic sequelae of deep venous thrombosis. Journal of vascular surgery. 1986 Jul 1;4(1):73-9.
7. Benotti JR, Dalen JE. The natural history of pulmonary embolism. Clinics in chest medicine. 1984 Sep 1;5(3):403-10.