

Design and Implementation of a Comprehensive Surveillance System for Venous Thromboembolism in a Defined Region Using Electronic and Manual Approaches

Thomas L. Ortel^{1,2} Katie Arnold³ Michele Beckman⁴ Audrey Brown⁵ Nimia Reyes⁴ Ibrahim Saber¹
Ryan Schulteis⁶ Bhavana Pendurthi Singh⁷ Andrea Sitlinger⁸ Elizabeth H. Thames¹

¹ Division of Hematology, Department of Medicine, Duke University Medical Center, Durham, North Carolina, United States

² Department of Pathology, Duke University Medical Center, Durham, North Carolina, United States

³ Duke Clinical Research Institute, Duke University School of Medicine, Durham, North Carolina, United States

⁴ Centers for Disease Control and Prevention, Atlanta, Georgia, United States

⁵ Social & Scientific Systems, Inc., Durham, North Carolina, United States

⁶ Durham Veterans' Administration Medical Center, Durham, North Carolina, United States

⁷ MedStar Georgetown University Hospital, Washington, District of Columbia, United States

⁸ Hematologic Malignancies and Cellular Therapy, Department of Medicine, Duke University Medical Center, Durham, North Carolina, United States

Address for correspondence Thomas L. Ortel, MD, PhD, Division of Hematology, Duke University Medical Center, Durham, NC 27710, United States (e-mail: thomas.ortel@duke.edu).

Appl Clin Inform 2019;10:552–562.

Abstract

Keywords

- ▶ venous thromboembolism
- ▶ deep vein thrombosis
- ▶ pulmonary embolism
- ▶ surveillance
- ▶ electronic medical record

Background Systematic surveillance for venous thromboembolism (VTE) in the United States has been recommended by several organizations. Despite adoption of electronic medical records (EMRs) by most health care providers and facilities, however, systematic surveillance for VTE is not available.

Objectives This article develops a comprehensive, population-based surveillance strategy for VTE in a defined geographical region.

Methods The primary surveillance strategy combined computerized searches of the EMR with a manual review of imaging data at the Duke University Health System in Durham County, North Carolina, United States. Different strategies of searching the EMR were explored. Consolidation of results with autopsy reports (nonsearchable in the EMR) and with results from the Durham Veterans' Administration Medical Center was performed to provide a comprehensive report of new VTE from the defined region over a 2-year timeframe.

Results Monthly searches of the primary EMR missed a significant number of patients with new VTE who were identified by a separate manual search of radiology records, apparently related to delays in data entry and coding into the EMR. Comprehensive searches incorporating a location-restricted strategy were incomplete due to the assigned residence reflecting the current address and not the address at the time of event. The most comprehensive strategy omitted the geographic restriction step and identified all patients with VTE followed by manual review of individual records to remove incorrect entries (e.g., outside the surveillance time period or geographic location; no evidence for VTE).

received
March 27, 2019
accepted after revision
June 16, 2019

© 2019 Georg Thieme Verlag KG
Stuttgart · New York

DOI <https://doi.org/10.1055/s-0039-1693711>.
ISSN 1869-0327.

Consolidation of results from the EMR searches with results from autopsy reports and the separate facility identified additional patients not diagnosed within the Duke system.

Conclusion We identified several challenges with implementing a comprehensive VTE surveillance program that could limit accuracy of the results. Improved electronic strategies are needed to cross-reference patients across multiple health systems and to minimize the need for manual review and confirmation of results.

Background and Significance

Deep vein thrombosis (DVT) and pulmonary embolism (PE), collectively referred to as venous thromboembolism (VTE), occurs in at least 350,000 individuals, and potentially up to 900,000 individuals, each year in the United States.¹⁻⁴ Hospitalization is a major risk factor for VTE,^{5,6} and many VTE events that occur in hospitalized patients could potentially be prevented.⁷ Case-fatality rates at 30 days are approximately 10% in all patients with new VTE, and this rate is increased by two- to threefold in patients with cancer.^{2,3} Although significant morbidity and mortality are associated with VTE, there is currently no systematic national surveillance system for this condition.

The Surgeon General of the United States issued a Call-to-Action to prevent DVT and PE in 2008,⁸ raising the question of whether a national surveillance program should be implemented. In 2010, the Centers for Disease Control and Prevention and the American Society of Hematology convened a national workshop of stakeholders on this issue, including relevant federal agencies, experts in the epidemiology and prevention of VTE, public health experts in VTE, and patient representatives.⁹ One of the conclusions from this workshop was that "...systematic surveillance of DVT and PE is needed to provide nationally representative data on the prevalence and annual incidence of DVT and PE in the U.S.," and that "...tracking and documenting changes in the incidence of DVT and PE through systematic surveillance will be important to enhance prevention efforts."⁹

Electronic medical records (EMRs) have been promoted by federal and state governments, and implemented by medical institutions, as a strategy to improve patient record-keeping and, thereby, clinical outcomes.¹⁰⁻¹² An optimal VTE surveillance strategy would utilize an institution's EMR to facilitate identification of patients with a new diagnosis of DVT and/or PE. Although most health care systems in the United States utilize an EMR, these systems do not necessarily share patient data, and patients frequently receive medical care at more than one medical facility. In addition, many search strategies use International Classification of Diseases, Ninth Revision (ICD-9), or Tenth Revision (ICD-10), codes, to identify patients with VTE. These codes have limitations that can complicate efforts to implement an accurate surveillance strategy for VTE in the United States.¹³⁻¹⁵

Objectives

The purpose of this study was to develop a comprehensive population-based surveillance strategy to identify new diag-

noses of VTE in residents within Durham County, North Carolina, United States. We combined electronic and manual search strategies to identify potential limitations of using an electronic strategy alone. This report describes the search strategies and performance characteristics for the surveillance program.

Methods

Location and Study Population

In 2012, the total population of Durham County, North Carolina, was approximately 280,000.¹⁶ Three hospitals are in the County: (1) Duke University Hospital (DUH), (2) Duke Regional Hospital (DRH), and (3) the Durham Veterans' Administration Medical Center (DURHAM VAMC). DUH and DRH are both part of Duke University Health System (DUHS) and use a shared EMR (Epic, Epic Systems Corporation; implemented at DUH in June 2013 and at DRH in March 2014). DUH is a tertiary and quaternary academic medical center with almost 1,000 inpatient beds, more than 40,000 admissions in fiscal year 2018, and more than 1,000,000 outpatient visits in fiscal year 2017. DRH is a community hospital in north Durham that has 369 inpatient beds, more than 16,000 admissions in fiscal year 2018, and more than 160,000 outpatient visits in fiscal year 2017. The DURHAM VAMC provides services to more than 200,000 Veterans living in central and eastern North Carolina, with 151 inpatient beds and a 100-bed Community Living Center. The DURHAM VAMC uses the Veterans Information Systems and Technology Architecture, an EMR that is integrated across the Veterans Health Administration (VHA) sites.¹⁷ The majority of the outpatient clinics and facilities located within Durham County are affiliated with DUH or DRH. The surveillance period began on April 1, 2012 and ended on March 31, 2014.

Data Sources

DUH/DRH EMR

Electronic data from clinical operations at DUH and DRH are archived at Duke Health Technology Solutions in Durham, North Carolina, and includes integrated clinical and financial data. Data are transferred daily from the source information through an ETL (extraction, transformation, and loading) process to house it in the Decision Support Repository (DSR).¹⁸ The DSR is a dimensionally modeled, standards-based database organized into multiple high-level subject areas, including demographics, encounters, provider orders, procedures, diagnoses, laboratory results, medications, vital signs, radiology, and pathology reports. Data, including ICD-9/10 and Current Procedural Terminology codes, are integrated from source systems to ensure consistency and quality and minimize

redundancy.¹⁹ Patient address information in the DSR is based on the most recent reported address, standardized using United States Postal Service standards, and assigned a location-specific geocode.²⁰

DUH/DRH Radiology Imaging Data

Electronic imaging data files at DUH and DRH are stored in a Picture Archiving and Communication System (PACS) maintained through DUHS.

Autopsy Reports

The DUH Autopsy Service performs approximately 350 autopsies annually. Autopsies are rarely performed at the DURHAM VAMC and were not included in this analysis. Autopsies are typically limited to the chest, abdomen, and pelvis; the limbs are not routinely examined unless requested. Autopsy reports were scanned and stored in the EMR, but were not electronically searchable at the time of our study. Autopsy records were matched to patients identified by other mechanisms during the surveillance period to avoid duplicate data entries.

DURHAM VAMC EMR

Clinical, administrative, and financial data for the VHA are archived in a relational corporate data warehouse (CDW), which includes patient demographic, physician demographic, clinical encounter, consult, order, pharmacy, laboratory, and radiology data. VHA autopsy data was not readily searchable at the time of our surveillance study.

Search Strategies

Electronic Searches of the DUHS EMR

The primary search strategy for the DUH/DRH EMR used the Duke Enterprise Data Unified Content Explorer (DEDUCE), a research portal developed to expedite access to clinical data stored in the DSR that has been available to Duke investigators since 2008.¹⁸ Searches using DEDUCE do not access the primary source data from the EMR. Three steps were used to restrict the data set to the selected cohort of interest:

1. The DSR data set was restricted to individuals living in Durham County, North Carolina, defined by geocoded address in the DSR.
2. Encounters that used an ICD-9 code consistent with a diagnosis of DVT and/or PE were identified. ICD-9 codes used for the searches are listed in **► Supplementary Table S1** (available in the online version). ICD-10 codes were not in use until after the surveillance period had closed.
3. The data set was further restricted to encounters occurring within the defined span of time for the study, either monthly or covering the entire surveillance period.

We explored several approaches to implementing these three steps (**► Fig. 1**):

1. Monthly searches, location-restricted: we used DEDUCE to query the DSR each month for individuals diagnosed

with a VTE, using the Durham County filter as determined by the most recent geocoded residence address provided by the patient. This approach was conducted monthly during the 2-year surveillance period, and for an additional 2 months after the surveillance period had ended to capture delayed entries into the EMR.

2. Single comprehensive search, location-restricted: we used DEDUCE to query the DSR approximately 1 year after the surveillance period had closed, using the Durham County filter as in the monthly searches. This approach was performed on two separate occasions, approximately 6 months apart.
3. Single comprehensive search, location-unrestricted: we used a direct structured query language query to probe the DSR to identify all patients with an ICD-9 code for DVT/PE covering the entire surveillance period. This approach was used to access an enterprise, encounter-keyed address table that included all addresses listed for each patient during the surveillance period. Addresses were compared with date of VTE diagnosis to specify the residence location at the time of diagnosis.

For each electronic search, individual patient records were manually reviewed to confirm the presence of VTE, the date of diagnosis, and the address at the time of the event. Criteria confirming the diagnosis included the presence of a new DVT and/or PE identified by appropriate imaging studies and documentation in the providers' notes concerning the diagnosis and plans for treatment.

Manual Review of DUH Radiology Imaging Data Identified through the EMR

As part of a separate program, all chest computed tomography angiography and Doppler ultrasound imaging data at DUH were reviewed on a daily basis by research staff in the Duke Hemostasis and Thrombosis Center to identify patients with a new DVT and/or PE. This review was independent from the surveillance program, but provided a list of patients with DVT and PE identified through manual review of the records to compare with records obtained through the use of DEDUCE and the DSR.

Manual Review of Duke Autopsy Reports

We reviewed all autopsies performed during the surveillance period, excluding individuals who did not live in Durham County at the time of death, still-born infants, infants surviving < 1 month from birth, and individuals for whom the autopsy was limited to examination of the brain only. Two investigators (I.S., T.L.O.) independently reviewed all reports to identify individuals with VTE identified at autopsy, and assessed the likelihood of whether the identified VTE contributed to the patient's death.

DURHAM VAMC EMR

The primary search strategy for the EMR records at DURHAM VAMC utilized the CDW, a strategy commonly used to identify cohorts of veteran patients for retrospective analysis.^{21,22}

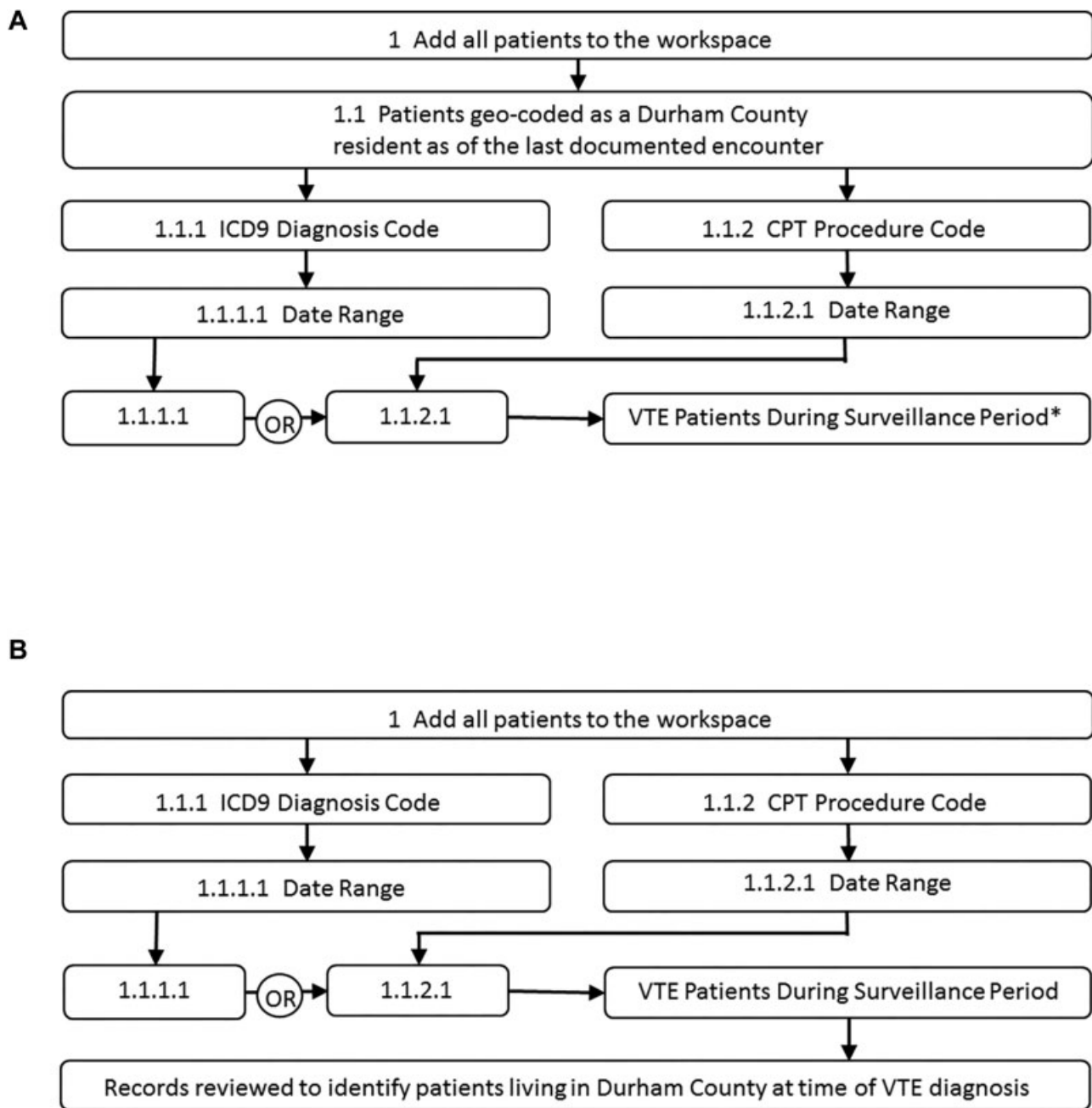


Fig. 1 Conceptual representation of query design in Duke Enterprise Data Unified Content Explorer (DEDUCE) Cohort Manager. (A) Strategy that uses a geo-spatially restricted subset of the Decision Support Repository (DSR). The Surveillance Period (identified by *) could be either monthly searches or a search covering the entire 2-year duration of the study. (B) Strategy that searches the full electronic medical record (EMR) to identify all patients with venous thromboembolism (VTE) diagnosed during the entire span of the study, with all addresses identified during the span of the study subsequently manually reviewed and correlated to the date of diagnosis with VTE.

Four steps were used to restrict the data set to the cohort of interest:

1. Limit to individuals having a radiology study that can diagnose VTE.
2. Limit to radiology studies having a diagnostic code consistent with the possibility of VTE (e.g., “Major Abnormality, Physician Aware,” “Critical Abnormality,” and “Significant Abnormality, Attention Needed”; we discussed this with the DURHAM VAMC Radiology Chief to ensure that 100% of

studies with findings of VTE are given a diagnostic code falling into one of these three categories).

3. Limit to radiology studies having occurred within the surveillance period.
4. Limit to individuals living in Durham County at the time of the event (as identified by CDW demographic data).

We searched the CDW for approximately 6 months after the surveillance period had closed, but covering the entire 2-year time span of the study. We extracted the patients’

Medical Record Number to confirm the radiologic diagnoses. Furthermore, we extracted the radiology study type, the date of the study, the county of residence (though limited to Durham County, this served as a control for validation purposes), and, finally, the radiology study report text.

Review and Cross-Reference of Individual Data Entries

Records of all individuals identified through the different strategies were reviewed by at least two separate members of the research team to confirm the diagnosis of new VTE. All confirmed individuals were assigned a unique patient identification number and entered into a REDCap database.²³ Each individual was cross-referenced across all data sources to confirm that each entry was unique. Duplicate entries were removed from the final data set.

Institutional Review Board Review

This population-based surveillance study was deemed exempt from review by the Duke Institutional Review Board since it met all of the specified criteria covered in 45 CFR 46.101 (b) 5.

Results

Searches Using the DUH/DRH EMR

Monthly Searches, Location-Restricted

We initially used DEDUCE to search the DSR monthly, with each search covering the span of time between the first and last days of each month (→Fig. 1A). This strategy identified 1,042 individuals during the 2-year surveillance period, but manual review found that almost half did not meet the study criteria (→Table 1). Common reasons for misidentified patients included: duplicate entries (patients identified in different months for the same event); VTE diagnosis predating the surveillance period; thrombosis involving a superficial rather than a deep vein; incorrect diagnosis (i.e., no VTE); and non-Durham County resident. After review, this strategy identified 540 unique patients living in Durham County with a new VTE during the surveillance period in the DUHS system (→Table 1).

In parallel with the monthly electronic searches, we identified 360 patients with DVT and/or PE who were living in Durham County through manual review of DUH PACS records. Comparing this list with the electronic surveillance data revealed that 116 of these patients had not been identified

by the monthly electronic searches, increasing the total number of cases of VTE in Durham County during the surveillance period to 656 (→Fig. 2A). There were no obvious reasons why these patients had not been captured through the monthly electronic searches, but this manual review identified that the monthly electronic searches were incomplete.

Two-Year Comprehensive Searches, Location Restricted

We next used the same strategy as the monthly searches but with the date range spanning the entire 2 years of the surveillance period (→Fig. 1A), and we ran the analysis approximately 1 year after the surveillance period was closed. This approach identified 792 patients living within Durham County with a new VTE during the study period (→Table 1). Multiple unique encounters for individual patients reflected use of the same ICD-9 codes during subsequent encounters with the same patient. Individuals not included in the final list of patients for this analysis had a VTE diagnosed outside the surveillance period, were living outside Durham County, did not have a VTE, or had superficial venous thrombosis only. The 792 patients identified by this search included: 503 of 540 patients identified by the monthly searches, 77 of 116 patients identified by the manual searches, and 212 additional patients who had not been previously identified. This observation confirmed our earlier concern that the monthly searches, as performed, were missing a significant proportion of patients diagnosed with new DVT/PE during the surveillance period. A subset of the patients missed by this search, but identified in the monthly searches, was no longer living in Durham County at the time of the surveillance run.

To evaluate reproducibility of this approach, we repeated the same strategy 6 months later. This analysis identified 798 patients who sustained a new VTE during the surveillance period and were living in Durham County at the time of diagnosis (→Table 1). Although this search identified a comparable total number of patients as the first search using this approach, there were differences in some of the individual patients identified. Seventy-eight patients identified in the first run of this search were not identified in the second run, and 84 identified in the second run had not been identified in the first run.

Two-Year Comprehensive Search, Location-Unrestricted

Because of concern that the geocoded address might be introducing errors into the search, we omitted the first step of the search strategy to comprehensively identify all

Table 1 Electronic searches of the DUH/DRH electronic medical record

| Search strategy | Unique encounters | Unique patients | Unique patients with VTE in Durham County |
|--|-------------------|-----------------|---|
| Monthly searches, location-restricted | NA ^a | 1,042 | 540 |
| Two-year comprehensive search, location-restricted (run 1) | 15,142 | 2,011 | 792 |
| Two-year comprehensive search, location-restricted (run 2) | 16,218 | 2,155 | 798 |
| Two-year comprehensive search, location-unrestricted | 91,284 | 9,935 | 978 |

Abbreviations: DRH, Duke Regional Hospital; DUH, Duke University Hospital; NA, not available; VTE, venous thromboembolism.

Note: For each search, the number of unique encounters, unique patients, and unique patients with VTE living in Durham County at the time of the event are provided.

^aIndividual patient encounters were not tabulated for the monthly searches.

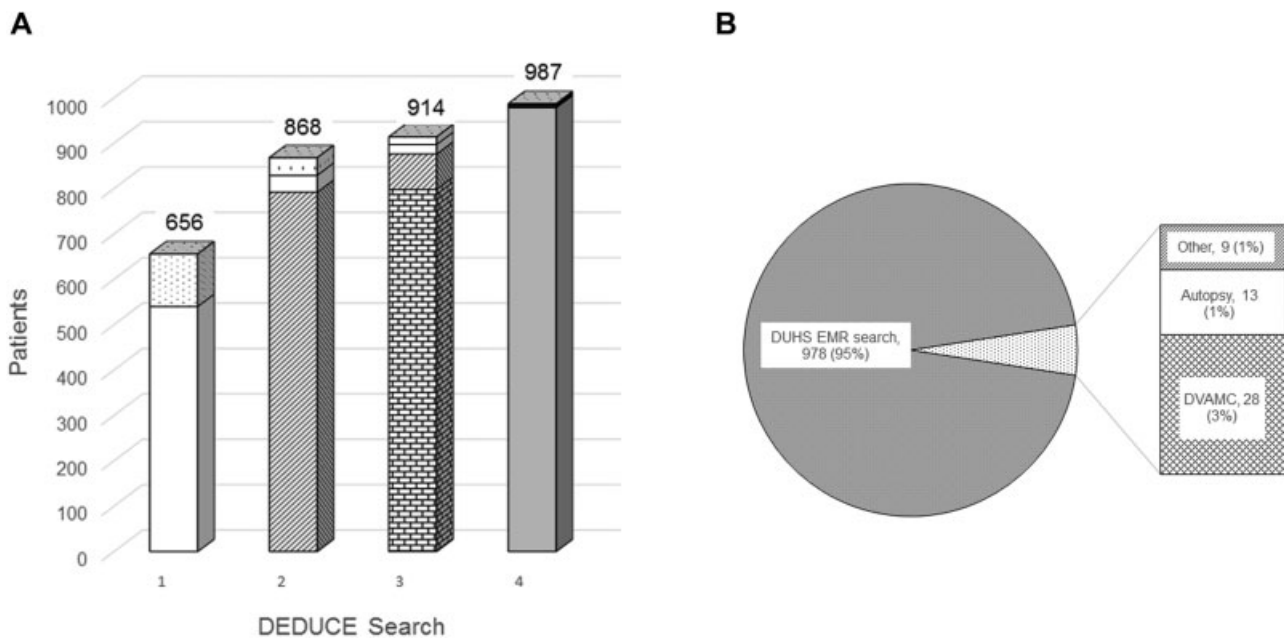


Fig. 2 Venous thromboembolism (VTE) surveillance results. (A) Aggregate results from four different searches of the Duke University Hospital (DUH)/Duke Regional Hospital (DRH) electronic medical record (EMR). Search 1: patients identified by location-restricted, monthly electronic searches of the Decision Support Repository (DSR) (white) with additional unique patients identified by manual review of DUH radiology reports (dots). Search 2: patients identified by a location-restricted, 2-year comprehensive electronic search of the DSR (diagonal lines) with additional patients identified by Search 1. Search 3: patients identified by a location-restricted, second 2-year comprehensive electronic search of the DSR (small bricks) with additional patients identified by Searches 1 and 2. Search 4: patients identified by a location-unrestricted, 2-year comprehensive search of the DSR with manual review of addresses and zip codes (gray) with additional patients identified by Searches 1, 2, and 3. The numbers at the top of each column represents the total number of unique patients identified by each strategy combined with unique patients from the previous results. Panel B includes a distribution of patients identified by all of the strategies employed as part of this study. The “Duke University Health System (DUHS) EMR search” represents the patients identified by the location unrestricted, 2-year comprehensive search. The column identifies the subsets of patients identified by: (1) the DUHS EMR using location-restricted strategies and manual review of records ($n = 9$); (2) the autopsy reports ($n = 13$); and (3) the Durham Veterans’ Administration Medical Center (DURHAM VAMC) search ($n = 28$).

patients with a new VTE during the surveillance period, regardless of residency (→ Fig. 1B). This query provided all addresses attached to the patient during the surveillance period, which were then manually reviewed to identify patients with new VTE who were living in Durham County at the time of diagnosis and during the surveillance period. This search strategy markedly increased the number of encounters and unique patients, but the increase primarily represented individuals who lived outside Durham County at the time of diagnosis (→ Table 1).

The location-unrestricted approach identified a total of 978 unique patients living in Durham County at the time of diagnosis with a new VTE (→ Table 1), including 78 additional patients who had not been identified by any of the previous search strategies. This method was most comprehensive and identified all but 9 patients identified by one (or more) of the prior electronic searches, as well as by manual review of the radiology reports (→ Fig. 2A). There was nothing different clinically about these 9 individuals that might explain why they were not identified by this strategy (→ Table 2). The three patients identified by manual review of the radiology reports only were hospitalized at the time of the event and had multiple comorbid conditions. It is possible that the ICD-9 codes used for the electronic searches were simply not used in any of these patients, although they did have imaging studies that documented the presence of VTE.

Manual Review of DUH Autopsy Reports

During the surveillance period, 677 autopsies were performed at DUH. After eliminating patients who did not live in Durham County at the time of death ($n = 551$) and autopsies performed on fetuses, infants less than 1 month of age, and those procedures limited to the brain only ($n = 29$), there were 97 patient reports for evaluation. Eighteen individuals had a VTE at autopsy, 12 with PE (12.4%) and 6 with DVT (6.2%). Five patients had been diagnosed with VTE prior to death and identified through one (or more) of the electronic searches, but 13 patients were identified by autopsy only (→ Fig. 2B).

Electronic Searches of the DURHAM VAMC EMR

During the surveillance period, there were 89,075 patient encounters at the DURHAM VAMC, and 6,848 of the veterans seen during these encounters lived in Durham County, representing approximately 2.5% of the Durham County population during the surveillance period. A total of 120 veterans from Durham County underwent imaging studies to evaluate for VTE, and 32 patients were diagnosed with new VTE during the surveillance period. Four of the 32 patients were also identified through electronic searches of the DUHS DSR, identifying a subset of veterans who sought medical care at a DUHS facility as well as the DURHAM VAMC.

Table 2 Patients not identified by comprehensive, location-unrestricted strategy

| Patient | Search strategy ^a | | | | VTE event | Diagnostic imaging |
|---------|------------------------------|---|---|---|----------------------------|----------------------------|
| | A | B | C | D | | |
| 1 | – | + | + | – | Portal vein thrombosis | CT abdomen |
| 2 | – | + | + | – | Splenic vein thrombosis | CT abdomen |
| 3 | + | – | – | – | Left subclavian thrombosis | Ultrasound |
| 4 | + | – | – | – | Bilateral PE | CT angiogram |
| 5 | – | – | – | + | LLE DVT | Ultrasound |
| 6 | – | – | – | + | RUE DVT | Ultrasound |
| 7 | – | – | – | + | RUE DVT | Ultrasound |
| 8 | + | – | – | – | LLE DVT | Ultrasound |
| 9 | + | – | + | – | PE | Ventilation/Perfusion scan |

Abbreviations: CT, computed tomography; DUHS, Duke University Health System; DVT, deep venous thrombosis; EMR, electronic medical record; LLE, left lower extremity; PE, pulmonary embolism; RUE, right upper extremity; VTE, venous thromboembolism.

Note: Nine patients identified by location-restricted searches of the DUHS EMR, or manual review of radiology reports, who were not identified by the location-unrestricted strategy.

^aThe search strategies were: A, Monthly electronic searches, location-restricted; B, First 2-year electronic search, location-restricted; C, Second 2-year electronic search, location-restricted; D, Manual review of radiology reports.

Consolidation of Surveillance Data

Consolidation of the results identified a total of 1,028 unique patients who lived in Durham County during the surveillance study period who sustained a new VTE (→Fig. 2B). The majority of individuals ($n = 1,000$) were identified through one or more of the searches conducted within the DUHS system (electronic and manual), consistent with the distribution of veteran and nonveteran individuals within Durham County. For individuals identified within the DUHS system, sensitivity was highest for the 2-year comprehensive EMR search that was location-unrestricted (→Table 3). Removing the location restriction resulted in a very low specificity for the target population, however, as DUH is a large referral center for patients living outside Durham County (→Table 3). Twenty-eight patients were identified through the DURHAM VAMC search only, representing approximately 3% of the total population in Durham County with VTE during the surveillance period (→Fig. 2B).

Given the estimated total population in Durham County during the surveillance period, the estimated annual frequency of VTE was approximately 1.8 events per 1,000 individuals (including incident events as well as recurrent events in individuals with a prior VTE), consistent with prior studies.^{1–3,7} The rate of new VTE diagnosis was relatively consistent on a month-by-month basis, with an average of 43 new diagnoses per month (→Fig. 3, solid line). If monthly searches of the DUHS EMR had been used alone, over 45% of all cases would not have been identified (→Fig. 3, dashed line).

Discussion

Public health surveillance is defined as “...the ongoing systematic collection, analysis, and interpretation of data, closely integrated with the dissemination of these data to public health practitioners, clinicians, and policy makers responsible for preventing and controlling disease and injury.”²⁴ Active

Table 3 Specificity and sensitivity of the different search strategies

| | Specificity ^a | Sensitivity ^b |
|--|--------------------------|--------------------------|
| Monthly EMR search, location-restricted | 52% (540/1,042) | 54% (540/1,000) |
| Two-year comprehensive EMR search 1, location-restricted | 39% (792/2,011) | 79% (792/1,000) |
| Two-year comprehensive EMR search 2, location-restricted | 37% (798/2,155) | 80% (798/1,000) |
| Two-year comprehensive EMR search, location-unrestricted | 10% (978/9,935) | 98% (978/1,000) |
| All DUHS searches (electronic and manual) | – | 97% (1,000/1,028) |

Abbreviations: DUHS, Duke University Health System; DURHAM VAMC, Durham Veterans' Administration Medical Center; DVT, deep venous thrombosis; EMR, electronic medical record; PE, pulmonary embolism.

^aFor the four EMR searches, specificity represents the number of “true positive” individuals (i.e., living in Durham County with new DVT/PE diagnosed during the surveillance period) over the total number of individuals identified by the specific search strategy.

^bFor the four EMR searches, sensitivity represents the number of “true positive” individuals (i.e., living in Durham County with new DVT/PE diagnosed during the surveillance period) identified by each search strategy over the total number of individuals identified by all DUHS strategies combined (electronic and manual searches, and autopsy reports). For the fifth search, sensitivity represents the total number of individuals identified by DUHS strategies divided by the total number of individuals identified by all approaches (DUHS and DURHAM VAMC).

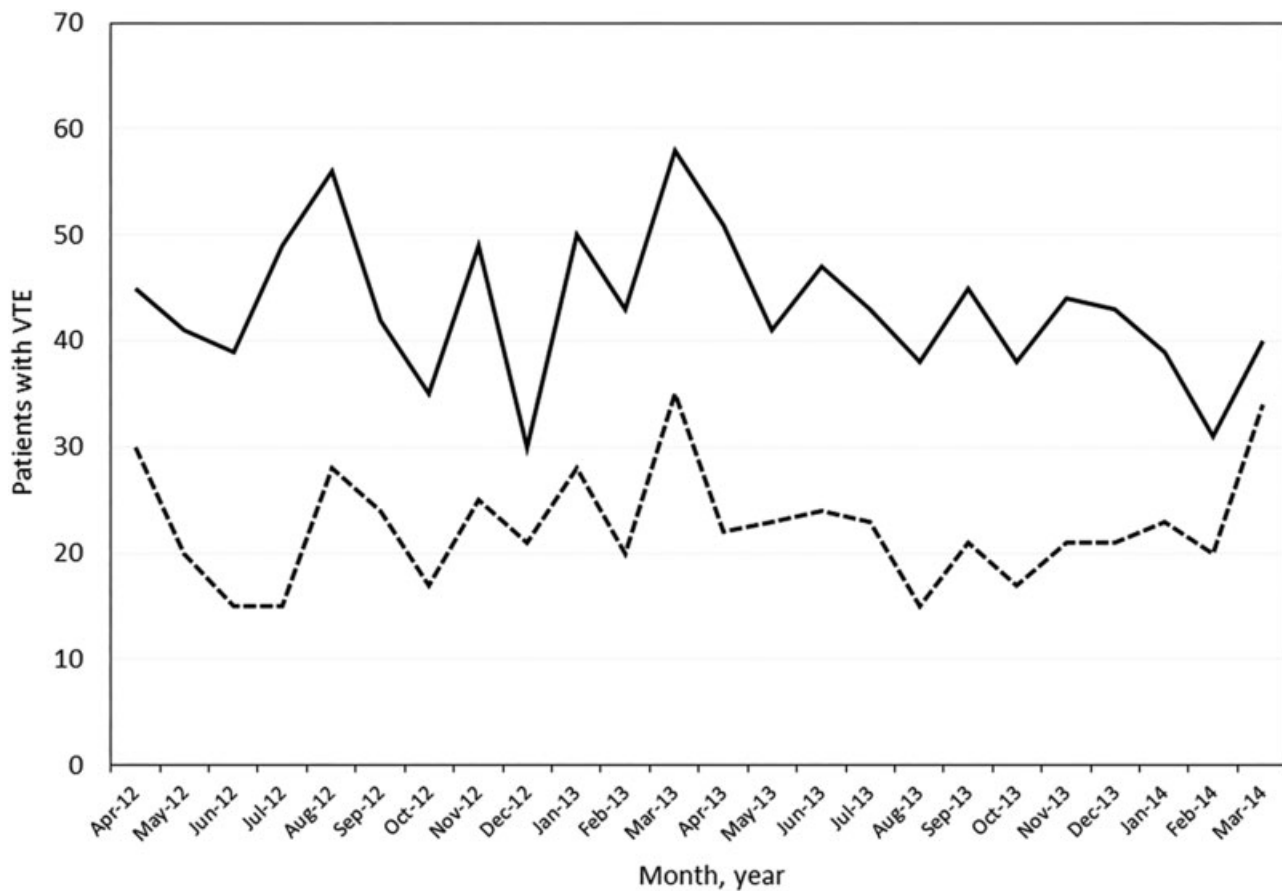


Fig. 3 New venous thromboembolism (VTE) diagnoses per month in Durham County. This figure includes all VTE diagnosed through the Duke University Hospital (DUH)/Duke Regional Hospital (DRH) electronic medical record (EMR), the manual review of autopsy reports performed at DUH, and through review of the Durham Veterans' Administration Medical Center (DURHAM VAMC) EMR (solid line) compared with VTE identified by the monthly searches of the DUH/DRH EMR only (dashed line).

surveillance programs exist for a variety of clinical conditions, including cancer,²⁵ respiratory illnesses,^{26,27} and multiple infectious diseases.^{28,29} Recent advances in surveillance strategies include electronic reporting of laboratory records from health care facilities to local and state public health agencies³⁰ as well as developing strategies to conduct real-time surveillance using electronic case reporting.³¹

Although governmental and regulatory agencies such as the Agency for Healthcare Research and Quality and the Centers for Medicare & Medicaid Services (CMS) have adopted VTE measures to assess hospital quality, most of these activities rely on the use of administrative codes, primarily using ICD-9 (or ICD-10) diagnosis codes.^{13,32} Concerns about the validity of these administrative codes have been raised, however, including accuracy of diagnosis in the outpatient setting¹⁴ and the risk of surveillance bias related to increased rates of VTE imaging.³³ Manual review of VTE cases identified by ICD-9 codes demonstrated frequent misclassification of preventable VTE,³⁴ raising questions about the validity of data collected automatically using the CMS Meaningful Use VTE-6 measure, a metric designed to use EMR data to identify patients with potentially preventable VTE who did not receive thromboprophylactic therapy between admission to the hospital and day prior to the diagnostic test order date.³⁵

We explored whether DEDUCE, a research portal developed at DUH to probe data derived from the EMR,^{18,20} could be used as a core strategy for implementing a VTE surveillance strategy across DUHS and/or Durham County, North Carolina. Although DUH and, to a lesser extent, the DURHAM VAMC, are large referral hospitals, we elected to restrict the patient population surveyed to Durham County to reflect a regional pattern in which most patients would be seeking care at the closest facility. To understand how comprehensive this strategy would be, we also conducted parallel searches of radiology reports and autopsy records at DUH, and ran a similar search for patients with VTE diagnosed at the DURHAM VAMC.

Performing monthly searches on the EMR was the least sensitive strategy (→Table 3). This may have reflected delays in entering diagnosis codes into the EMR, so encounters occurring late in the month, or patients with prolonged hospitalizations, may not have been captured, and subsequent searches did not rescan prior months to capture potentially missed encounters. In contrast, running a search spanning the full surveillance period after the study period had closed substantially improved sensitivity of the search (→Table 3). A limitation of this approach was that these searches used the DSR, and the geocoded address was the most recent address, not necessarily the address at the

time of VTE diagnosis. This limitation would have less of an impact on regular, monthly searches, but would result in misidentification of those patients who had relocated in the interval between VTE diagnosis and electronic search.

The location-unrestricted search strategy was the most comprehensive, since it identified all patients diagnosed with VTE during the study period and provided all addresses reported by the patient during the 2-year surveillance period. Sensitivity was 98% (→Table 3), but the total number of unique patients that had to be manually reviewed to confirm the diagnosis and location of residency was almost five times greater than any of the other searches due to the identification of patients diagnosed with VTE at DUHS who lived outside Durham County at the time of diagnosis (→Table 1). This increased the time and manual effort required to perform the surveillance, limiting the benefit of an approach built on an automated electronic search strategy.

It has been estimated that as many as one-quarter of patients experiencing PE present with sudden death.³⁶ The prevalence of acute PE at autopsy has been reported to be 12 to 14%.^{37,38} We identified PE in 12.4% of autopsies from Durham County during the surveillance period, and a DVT without PE in an additional 6.2% of autopsies. These data suggest that there continues to be a subset of patients with PE who present with sudden death, which needs to be taken into consideration when estimating the true burden of disease with VTE. Limitations to incorporating autopsy reports include the small number of autopsies performed, and the fact that autopsies are frequently limited to the chest, abdomen, and/or pelvis, which would not identify DVT of the limbs.

Another difficulty with performing VTE surveillance is that most communities have more than one health care facility, which can complicate efforts to develop more extensive surveillance strategies across large communities.³⁹ For example, 4 of the 32 patients with a new VTE diagnosed at the DURHAM VAMC were also identified by surveillance at DUHS and would not have been identified as duplicate encounters if the surveillance data had not been consolidated and compared. Conversely, 28 patients diagnosed at the DURHAM VAMC would not have been captured in the final count if we had not also conducted this search at the same time (→Fig. 2B). In addition, some individuals who live in Durham County obtain their health care with providers and facilities outside the county, which would result in an underestimate of the overall incidence in a geographically limited region. These issues will need to be taken into consideration when expanding a VTE surveillance strategy to include larger and more populated regions of the country.

Our study identified several additional difficulties that will be encountered in any effort to implement a national, or even statewide, surveillance strategy for VTE. First, the EMR at DUH and DRH transitioned from a legacy system to Epic during the surveillance period, creating the potential to disrupt active surveillance due to the introduction of new information technology systems. Hospitals and health care systems are frequently updating software, and surveillance strategies would need to be able to adapt to these changes. Second, the introduction of ICD-10 codes occurred after our

surveillance period was complete, and while this change would not impact on the approach to surveillance using an EMR, future studies would need to include current diagnosis and procedure codes. Third, while electronic strategies will be critical to implement a VTE surveillance program, manual oversight will continue to be necessary to minimize errors in the data collected. Incorporation of a natural language processing strategy to review radiology reports for VTE, which has been described in several reports,^{40–43} could be used to complement our strategy and decrease the number of records that need manual review. Last, linking imaging data to therapeutic interventions will be essential to assess the clinical impact of strategies introduced to prevent VTE, to monitor changes in incidence over time.

Although there will be difficulties with implementing a VTE surveillance strategy that integrates data across multiple institutions, we feel that this is a universal problem that needs to be addressed. Proposed uses for a VTE surveillance tool would include accurate quantitation of VTE rates, identification of best practices to prevent VTE, determination of VTE rates over time to assess impact of interventions, and improving outcomes for patients. In addition, linking the VTE surveillance data with laboratory results and/or medication data would provide insights that could enhance diagnostic strategies and therapeutic interventions. This strategy could be implemented in any health care system that uses an EMR.

Conclusion

We have shown that a VTE surveillance program incorporating an electronic search strategy of the EMR is feasible, but still requires careful review of the data collected electronically to confirm new cases as well as remove duplicate entries. Implementation of a national surveillance program for VTE will require taking into account limitations of using ICD-9 (now ICD-10) diagnosis codes, geographic mobility of the population, and differences between EMR programs used by health care systems.

Clinical Relevance Statement

Venous thromboembolism is a life-threatening event that can potentially be prevented in certain high-risk patient populations. There is currently no systematic national surveillance system for venous thromboembolism in place. Strategies using electronic searches of electronic medical records provide one approach to addressing this issue, but these searches may still require considerable manual oversight to provide accurate, comprehensive information.

Multiple Choice Questions

1. Limitations of using an electronic search strategy to identify patients with venous thromboembolism (VTE) include:
 - a. Duplicate identification of individual patients.
 - b. Incorrect identification of patients not having a VTE.

- c. Lack of identification of patients diagnosed at a different institution.
- d. All of the above.

Correct Answer: The correct answer is option d, all of the above. Duplicate identification may occur in patients repetitively identified at the same or different institutions. Incorrect identification may occur when incorrect ICD-9 (or ICD-10) codes are entered into the patient record. Lack of identification may occur when a patient is diagnosed with VTE at a different institution that does not share access to electronic medical records.

2. An important reason for establishing a national surveillance system for VTE is:
 - a. VTE is a rare condition that is poorly understood.
 - b. Tracking unprovoked VTE is an accepted measure for assessing hospital quality.
 - c. Surveillance will better define VTE incidence and impact of preventive strategies.
 - d. All of the above.

Correct Answer: The correct answer is option c, surveillance will better define disease incidence and impact of preventive strategies. VTE is not rare, but the exact incidence has not been clearly defined. Hospital-associated VTE, which is considered provoked, is used as a quality measure for hospitals. A national surveillance strategy could facilitate comparison between different preventive strategies by providing data before and after implementation of different thromboprophylactic strategies.

Protection of Human and Animal Subjects

This population-based surveillance study was deemed exempt from review by the Duke IRB since it met all of the specified criteria covered in 45 CFR 46.101 (b) 5.

Conflict of Interest

T.L.O. reports grants from Centers for Disease Control & Prevention, during the conduct of the study. R.S. reports personal fees from Boehringer Ingelheim, outside the submitted work.

Acknowledgments

We wish to thank Kristin Hudd-Byrne, Kristen Bagby, RN, and Sheila Lambert-Adams for assistance with manual review of the DUH/DRH EMR reports. We also thank Alan Proia, MD, PhD, and Brenda Dudzinski for their assistance with review of the Duke autopsy reports. This article was supported by Grant/Cooperative Agreement U50DD000897 from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

References

- 1 Heit JA. The epidemiology of venous thromboembolism in the community. *Arterioscler Thromb Vasc Biol* 2008;28(03):370–372
- 2 Cushman M, Tsai AW, White RH, et al. Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. *Am J Med* 2004;117(01):19–25
- 3 Puurunen MK, Gona PN, Larson MG, Murabito JM, Magnani JW, O'Donnell CJ. Epidemiology of venous thromboembolism in the Framingham Heart Study. *Thromb Res* 2016;145:27–33
- 4 Yusuf HR, Tsai J, Atrash HK, et al; Centers for Disease Control and Prevention (CDC). Venous thromboembolism in adult hospitalizations - United States, 2007-2009. *MMWR Morb Mortal Wkly Rep* 2012;61(22):401–404
- 5 Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ III. Risk factors for deep vein thrombosis and pulmonary embolism: a population-based case-control study. *Arch Intern Med* 2000;160(06):809–815
- 6 Tsai J, Grant AM, Beckman MG, Grosse SD, Yusuf HR, Richardson LC. Determinants of venous thromboembolism among hospitalizations of US adults: a multilevel analysis. *PLoS One* 2015;10(04):e0123842
- 7 Beckman MG, Hooper WC, Critchley SE, Ortel TL. Venous thromboembolism: a public health concern. *Am J Prev Med* 2010;38(4, Suppl):S495–S501
- 8 Office of the Surgeon General (US); National Heart, Lung, and Blood Institute (US). The Surgeon General's Call to Action to Prevent Deep Vein Thrombosis and Pulmonary Embolism. Rockville, MD: Office of the Surgeon General (US); 2008. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK44178/>. Accessed July 17, 2019
- 9 Raskob GE, Silverstein R, Bratzler DW, Heit JA, White RH. Surveillance for deep vein thrombosis and pulmonary embolism: recommendations from a national workshop. *Am J Prev Med* 2010;38(4, Suppl):S502–S509
- 10 Steinbrook R. Health care and the American Recovery and Reinvestment Act. *N Engl J Med* 2009;360(11):1057–1060
- 11 Adler-Milstein J, DesRoches CM, Kralovec P, et al. Electronic health record adoption in US hospitals: progress continues, but challenges persist. *Health Aff (Millwood)* 2015;34(12):2174–2180
- 12 Adler-Milstein J, Furukawa MF, King J, Jha AK. Early results from the hospital Electronic Health Record Incentive Programs. *Am J Manag Care* 2013;19(07):e273–e284
- 13 Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying venous thromboembolism using administrative and claims data. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl 1):154–162
- 14 Fang MC, Fan D, Sung SH, 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–e143
- 15 White RH, Garcia M, Sadeghi B, et al. Evaluation of the predictive value of ICD-9-CM coded administrative data for venous thromboembolism in the United States. *Thromb Res* 2010;126(01):61–67
- 16 USC Bureau. Secondary. Available at: <https://www.census.gov>. Accessed July 5, 2019
- 17 Byrne CM, Mercincavage LM, Pan EC, Vincent AG, Johnston DS, Middleton B. The value from investments in health information technology at the U.S. Department of Veterans Affairs. *Health Aff (Millwood)* 2010;29(04):629–638
- 18 Horvath MM, Winfield S, Evans S, Slopek S, Shang H, Ferranti J. The DEDUCE Guided Query tool: providing simplified access to clinical data for research and quality improvement. *J Biomed Inform* 2011;44(02):266–276
- 19 Horvath MM, Ruscovitch SA, Brinson S, Shang HC, Evans S, Ferranti JM. Modular design, application architecture, and usage of a self-service model for enterprise data delivery: the Duke Enterprise Data Unified Content Explorer (DEDUCE). *J Biomed Inform* 2014;52:231–242
- 20 Anderson DJ, Rojas LF, Watson S, et al; CDC Prevention Epicenters Program. Identification of novel risk factors for community-acquired *Clostridium difficile* infection using spatial statistics

- and geographic information system analyses. *PLoS One* 2017;12(05):e0176285
- 21 Patel VA, Romain PS, Sanchez J, Fisher DA, Schulte RD. Obstructive sleep apnea increases the risk of cardiopulmonary adverse events associated with ambulatory colonoscopy independent of body mass index. *Dig Dis Sci* 2017;62(10):2834–2839
 - 22 Chung MJ, Hansen JD, Schulte RD, et al. Understanding operator stent choice in the catheterization laboratory using a pre-procedure survey: opportunities for quality improvement. *Cardiovasc Revasc Med* 2017;18(08):588–591
 - 23 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42(02):377–381
 - 24 Richards CL, Iademarco MF, Atkinson D, et al. Advances in public health surveillance and information dissemination at the Centers for Disease Control and Prevention. *Public Health Rep* 2017;132(04):403–410
 - 25 Noone AM, Cronin KA, Altekruse SF, et al. Cancer incidence and survival trends by subtype using data from the surveillance epidemiology and end results program, 1992–2013. *Cancer Epidemiol Biomarkers Prev* 2017;26(04):632–641
 - 26 Moorman JE, Akinbami LJ, Bailey CM, et al. National surveillance of asthma: United States, 2001–2010. *Vital Health Stat* 3 2012; (35):1–58
 - 27 Ford ES, Croft JB, Mannino DM, Wheaton AG, Zhang X, Giles WH. COPD surveillance—United States, 1999–2011. *Chest* 2013;144(01):284–305
 - 28 Dugan VG, Blanton L, Elal AIA, et al. Update: influenza activity - United States, October 1–November 25, 2017. *MMWR Morb Mortal Wkly Rep* 2017;66(48):1318–1326
 - 29 Childs JE, Krebs JW, Real LA, Gordon ER. Animal-based national surveillance for zoonotic disease: quality, limitations, and implications of a model system for monitoring rabies. *Prev Vet Med* 2007; 78(3–4):246–261
 - 30 Lamb E, Satre J, Hurd-Kundet G, et al; Centers for Disease Control and Prevention (CDC). Update on progress in electronic reporting of laboratory results to public health agencies - United States, 2014. *Morb Mortal Wkly Rep* 2015;64(12):328–330
 - 31 Mac Kenzie WR, Davidson AJ, Wiesenthal A, et al. The promise of electronic case reporting. *Public Health Rep* 2016;131(06):742–746
 - 32 Yusuf HR, Reyes N, Zhang QC, Okoroh EM, Siddiqi AE, Tsai J. Hospitalizations of adults ≥ 60 years of age with venous thromboembolism. *Clin Appl Thromb Hemost* 2014;20(02):136–142
 - 33 Bilimoria KY, Chung J, Ju MH, et al. Evaluation of surveillance bias and the validity of the venous thromboembolism quality measure. *JAMA* 2013;310(14):1482–1489
 - 34 Lau BD, Haut ER, Hobson DB, et al. ICD-9 code-based venous thromboembolism performance targets fail to measure up. *Am J Med Qual* 2016;31(05):448–453
 - 35 Farrow NE, Lau BD, John Bull EA, et al. Is the Meaningful Use Venous Thromboembolism VTE-6 Measure Meaningful? A retrospective analysis of one hospital's VTE-6 cases. *Jt Comm J Qual Patient Saf* 2016;42(09):410–416
 - 36 Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ III. Predictors of survival after deep vein thrombosis and pulmonary embolism: a population-based, cohort study. *Arch Intern Med* 1999;159(05):445–453
 - 37 Goldhaber SZ, Hennekens CH, Evans DA, Newton EC, Godleski JJ. Factors associated with correct antemortem diagnosis of major pulmonary embolism. *Am J Med* 1982;73(06):822–826
 - 38 Stein PD, Henry JW. Prevalence of acute pulmonary embolism among patients in a general hospital and at autopsy. *Chest* 1995; 108(04):978–981
 - 39 Wendelboe AM, Campbell J, McCumber M, et al. The design and implementation of a new surveillance system for venous thromboembolism using combined active and passive methods. *Am Heart J* 2015;170(03):447–54.e18
 - 40 Tian Z, Sun S, Eguale T, Rochefort CM. Automated extraction of VTE events from narrative radiology reports in electronic health records: a validation study. *Med Care* 2017;55(10):e73–e80
 - 41 McPeck Hinz ER, Bastarache L, Denny JC. A natural language processing algorithm to define a venous thromboembolism phenotype. *AMIA Annu Symp Proc* 2013;2013:975–983
 - 42 Gálvez JA, Pappas JM, Ahumada L, et al. The use of natural language processing on pediatric diagnostic radiology reports in the electronic health record to identify deep venous thrombosis in children. *J Thromb Thrombolysis* 2017;44(03):281–290
 - 43 Nelson RE, Grosse SD, Waitzman NJ, et al. Using multiple sources of data for surveillance of postoperative venous thromboembolism among surgical patients treated in Department of Veterans Affairs hospitals, 2005–2010. *Thromb Res* 2015;135(04):636–642