Predicting Dynamically Evolving New-Onset Venous Thromboembolic (VTE) Event Risk in Hospitalized Patients.

by

Tiffany Purcell Pellathy

Bachelor of Science of Nursing, The Johns Hopkins University School of Nursing, 2001

Master of Science, Georgetown University School of Nursing and Health Studies, 2006

Submitted to the Graduate Faculty of the

School of Nursing in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF NURSING

This dissertation was presented

by

Tiffany Purcell Pellathy

It was defended on

December 9, 2020

and approved by

Salah Al-Zaiti, PhD, RN, ANP-BC, FAHA, Associate Professor, University of Pittsburgh School of Nursing, Department of Acute and Tertiary Care

Gilles Clermont, MD, MSc, Professor, University of Pittsburgh School of Medicine, Department of Critical Care Medicine

Artur Dubrawski, PhD, Research Professor, Carnegie Mellon University, School of Computer Science

Young Ji Lee, PhD, MS, RN, Assistant Professor, University of Pittsburgh School of Nursing, Department of Acute and Tertiary Care

Michael R. Pinsky, MD, CM, Dr hc, FCCP, MCCM, Professor, University of Pittsburgh School of Medicine, Department of Critical Care Medicine

Melissa Saul, MS, Adjunct Assistant Professor, Health Information Management, University of Pittsburgh School of Health and Rehabilitation Sciences

Thesis Advisor/Dissertation Director:
Marilyn Hravnak, PhD, RN, ACNP-BC, CCNS, MCCM, FAAN, Professor, University of Pittsburgh School of Nursing, Department of Acute and Tertiary Care

Predicting Dynamically Evolving New-Onset Venous Thromboembolic (VTE) Event Risk in Hospitalized Patients

Tiffany Purcell Pellathy

University of Pittsburgh, 2020

Word count 345/350

Background. Hospital acquired (HA) venous thromboembolism (VTE) is the leading cause of preventable hospital death. VTE pathology and symptoms evolve slowly over hours to days. No current VTE risk assessment models incorporate the progressive accrual of dynamic patient data over time of hospitalization. Classification algorithms which incorporate prediction time windows hold promise for closing this gap.

Methods. An observational, retrospective, cohort study was conducted to develop predictive models to classify patients (n=2370) at risk for HA-VTE during SDU admission. Binary logistic regression (BLR), naïve Bayes (NB), Random Forest (RF), and Gradient Boosted Decision Tree (GBDT) algorithms were used to train models for two prediction time windows. Performance was evaluated with 10-fold stratified cross-validation. Models (S+/-) were developed to differentiate patients suspected of HA-VTE who underwent diagnostic radiology evaluation (n=760) from those not suspected/not tested (n=1614). A second set of models (C+/-) were then developed to differentiate between confirmed positive (n-47) and negative (n-713) diagnostic test results. Models were built using a stage-wise process that increased data granularity with each stage: 1) present-on-admission data; 2) low frequency (LF) medication and laboratory data added; and 3) addition of high frequency (HF) vital sign data, collected at a rate of every 20 seconds.

Performance was evaluated at each stage using metrics robust to class imbalance and prioritizing recall (TPR).

Results. All models demonstrated improved precision-recall performance with progressive addition of dynamic clinical data. Using dynamic LF and HF data, at a prediction time 24 hours in advance of HA-VTE event, the S+/- NB model TPR was 76% (AUPRC .52, PPV 46%, AUROC .60) and RF and GBDT models identified true negatives with a specificity of 80%, and the C+/- NB model had a 91% TPR (AUPRC .77, PPV 53%, AUROC .68). Dynamic hematologic labs, BP, HR, and RR values were identified as important predictors of HA-VTE event outcomes, with importance varying by time prediction window.

Conclusion. Classification algorithms applied to routinely collected dynamic clinical data can produce models with improved HA-VTE risk prediction ability over static data models and have the potential to improve detection of at-risk patients.

Table of Contents

List of Tables

List of Figures

.

1.0 Dissertation Proposal

Section 1.0 of this document details the dissertation proposal that was presented and approved at the comprehensive examination and overview.

1.1 Specific Aims

Failure to rescue (FTR), a nurse sensitive national metric of care quality, refers to the death of a hospitalized patient from a treatable complication, and is potentiated by failure to recognize and appropriately respond to early signs and symptoms of complications. [1-5] There is a paucity of research examining patient features predictive of FTR complications. [6] Such information could shift the current paradigm of nursing surveillance to earlier recognition, prevention and treatment of FTR complications, thereby saving lives. Hospital acquired venous thromboembolism (HA-VTE), a FTR complication manifesting as deep vein thrombosis (DVT) or pulmonary embolism (PE), is the leading cause of preventable hospital death, carrying a high mortality risk and a national cost burden of $7 billion annually.[7-9] Clinical signs and symptoms of evolving venous thromboembolism (VTE) are subtle, presenting gradually over hours to days. Current VTE risk assessment models, the cornerstone of prevention, have limited utility due to their complexity, lack of reliability, generalizability and external validation, and dependency on static data. [10, 11] Importantly, a critical gap in VTE risk modeling research is that while VTE risk and pathology evolves over the course of hospitalization, no current models incorporate the progressive accrual of dynamic patient data and pattern evolution over time in their modeling approaches.

Use of electronic health record (EHR) data could be leveraged to develop VTE risk models. The totality of routinely collected EHR data is massive in terms of volume, variety, and production at a rapid velocity in real-time. [12] Such big data could be used in machine learning (ML) analytic approaches to process time series clinical data to identify subtle, evolving feature patterns predictive of new-onset HA-VTE and address this gap.

To address the aforementioned gaps, we propose to assemble a large scale, multi-source, multi-dimensional VTE dataset, and in tandem, systematically define the EHR data elements associated with a new-onset HA-VTE diagnosis for ground truth case ascertainment and annotation. We will then apply machine learning approaches to develop models identifying feature patterns predictive of dynamically evolving new-onset HA-VTE in adult hospitalized patients. Specifically, we aim to:

1. **Assemble a large scale, multi-source, multidimensional dataset from adult hospital step down unit (SDU) patients.** Demographic and clinical baseline data will be linked to extracted high frequency (every 20 seconds) vital sign (VS) and lower frequency EHR data (medications, laboratory and diagnostic tests) from >3,000 patients from a Step-Down Unit census. These data will be linked, annotated and curated to comprise the data set for ML training, validation and testing. [13]
2. **Ground truth case ascertainment of new onset, HA-VTE.** Gold standard clinical case definition, subtype identification, and date and time of VTE diagnostic test will be based on natural language processing (NLP) and expert manual review identification of ground truth diagnostic tests and diagnostic codes. Associated data elements, derived from standardized coding systems, reflective of diagnostic processes and treatment decisions associated with HA-VTE will be identified during this process.

3. **Develop and evaluate models to predict new-onset HA-VTE risk in hospitalized patients.**

3a. Apply ML algorithms to develop and evaluate models for new-onset HA-VTE risk over time. A suite of ML classification algorithms will be employed to train models to predict new-onset HA-VTE risk over the time course of hospitalization, at various lead times before diagnosis.

3b. Compare the various models' capability to predict dynamically evolving HA-VTE. The models' predictive capability will be assessed and ranked by standard metrics (sensitivity, specificity, prediction values), area under the receiver operating curve (AUROC) and area under the precision-recall curve (AUPRC) at various fixed time windows prior to new-onset HA-VTE diagnosis.

This proposal aligns with the National Institute of Nursing Research's (NINR) strategic vision for nurse scientists to employ new strategies for collecting and analyzing multi-dimensional data sets to permit better understanding of the biological underpinnings of health and improve ways nurses prevent and manage illness. This innovative study coupled with the candidate's individualized training plan, under a strong and well-established team, represents initial steps in the applicant's research trajectory. This future line of inquiry will focus on using data science approaches to predict FTR complications risk, and develop, implement and test dynamic risk assessment models (RAM) to inform targeted prevention and treatment decisions. This research trajectory has the potential to improve nurse sensitive patient outcomes through the discovery of new knowledge to guide nursing surveillance practices (needed frequency of monitoring, staff allocation), clinical decision making (timely and accurate recognition, diagnosis, treatment selection), and care delivery systems (patient triage, diagnostic testing, adverse event prevention).

1.2 Background, Significance, and Innovation

1.2.1 Background

1.2.1.1 Failure to Rescue (FTR)

FTR represents the death of a hospitalized patient due to a treatable condition arising after admission and is a measure of hospital care quality incorporating institutional attributes, nursing surveillance and patient features. Landmark studies by Silber et al., Needleman et al., and Aiken et al., led to the recognition of 15 specific FTR complications and highlighted the complexity of identifying patients experiencing a clinically important deterioration. [1-3, 5, 14, 15] Five FTR complications (pneumonia, cardiac arrest/shock, upper gastrointestinal bleed, sepsis, and HA-VTE) are established as being nurse-sensitive, meaning the intervention and rescue of patients experiencing these complications is directly associated with nursing surveillance practice and their early identification of changes in patient condition. [16-18] (see Figure 1)

Figure 1. Hospital Organization, Nursing Organization and Patient Outcomes

Reprinted from Nursing Outlook, Vol 50/5, Aiken L, Clarke S, Sloane D, Hospital staffing, organization, and quality of care: Cross-national findings, Pages 187-194, (2002), with permission from Elsevier, License No. 4934311399533. DOI: 10.1067/mno.202.126696

Strategies focused on improving health care organization resources, workflow and staffing practices have been unsuccessful in reducing FTR rates.[19, 20] Additionally, FTR has evolved in the literature from its origin as a quality outcome measure to include clinical characteristics identified as nurse sensitive indicators that represent the nursing processes associated with FTR outcomes: failure to recognize, failure to escalate, and inadequate decision-making.[15, 21] FTR nurse sensitive indicators formally recognize the fact that nurses spend the majority of their time at patients' bedsides, and the impact that nursing's timely recognition of, escalation of, and appropriate intervention based on subtle changes indicative of clinical deterioration has on patient safety. The first step in the process, failure to recognize, is established as a multifaceted challenge. There is a paucity of research examining *patient features* predictive of FTR complications.[6, 22-24] Early identification of patient features associated with FTR risk and complication evolution could inform targeted prevention and early intervention. (see Figure 2)

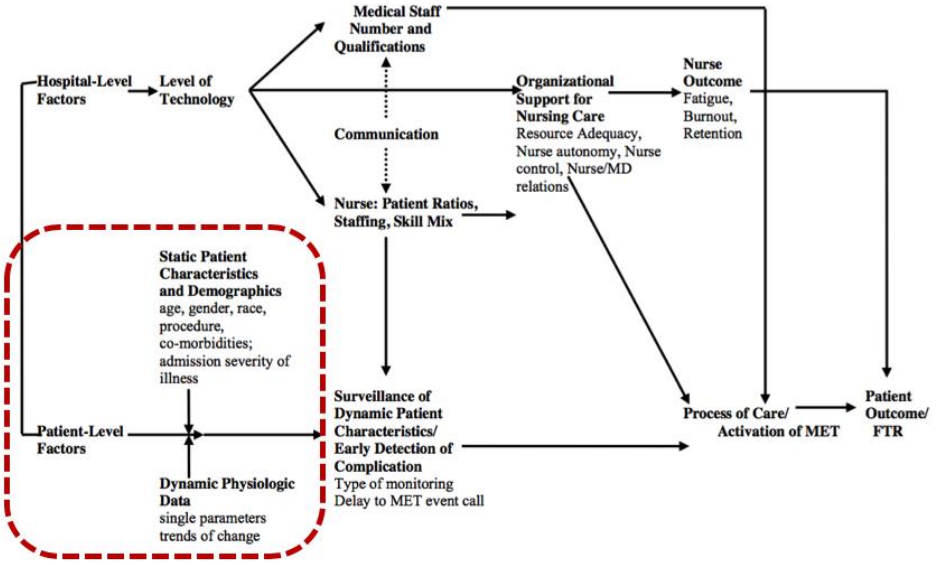


Figure 2. Conceptual Model for Causes of FTR: Incorporating Medical Emegency Team Availability

1.2.1.2 Inadequacy of General Scoring Systems in Reducing FTR Rates

A variety of scoring systems are available to support inpatient clinician decision making. Outcome risk prediction (i.e., APACHE) and organ dysfunction (i.e., SOFA, MODS) severity scoring systems have been considered routine decision support tools in patient care for the past three decades. Outside of the intensive care unit (ICU), scoring systems such as the modified early warning score (MEWS), the national early warning score (NEWS), and the quick sequential organ failure assessment (qSOFA) score [25-27] are tools that assist nurses with the detection of general physiological changes that may be indicative of early deterioration. The development of these risk prediction models, the variables included, and the indications for their use are detailed in Dissertation Manuscript #1 (Appendix A). Often used in conjunction with activation of an outreach or rapid response team, these systems are clinically useful, however their impact on FTR outcomes has been marginal.[28-31] This may be, in part, due to the fact that pattern recognition requires accounting for multiple interactive relationships across time among assessment variables. Pattern recognition, hypothesized to be central to clinician critical thinking, and foundational to recognizing and interpreting subtle clinical changes, [32, 33] is not accounted for in these systems.

1.2.1.3 Nursing Surveillance: A Prerequisite to Timely Intervention

Nursing surveillance is a nursing intervention critical to patient and safety outcomes. Nurses are the hospital surveillance network for the early detection of adverse occurrences, complications and errors.[20] Nursing surveillance involves the purposeful and ongoing acquisition, interpretation and synthesis of patient data to inform clinical decision-making.[17] Surveillance

requires clinician vigilance: the ability to place obtained data within the context of one's own knowledge, experience and education to inform pattern recognition, and risk calculation and readiness to act. Vigilance and surveillance comprise the critical thinking skills prerequisite to informed and timely nursing action. Nursing surveillance is affected by a variety of factors: nurse education level, clinician experience, clinician fatigue, alarm fatigue, workplace processes, staffing ratios and unit skill mix. Notably, inadequate surveillance is strongly associated with increased FTR rates. [5, 34, 35] Identification of feature patterns predictive of FTR complications can objectively support clinician pattern recognition essential for quality nursing surveillance and timely nursing intervention.

1.2.1.4 HA-VTE as a FTR Complication: Complexity and Persistence

Hospital acquired VTE can present as either deep vein thrombosis (DVT) or pulmonary embolism (PE). While DVT is the more common manifestation of this disease, PE is a more serious complication, associated with a higher rate of mortality.[36, 37] The nurse sensitive FTR complication of HA-VTE is the leading cause of preventable hospital death carrying a high mortality risk and a national cost burden of $7 billion annually.[7-9] Additionally, VTE is the third most common cause of cardiovascular death in the US, accounting for 2 million new diagnoses each year.[38, 39] Reasons cited for the persistence of this major public health problem include: 1) complexity of known VTE risk factors and unknown gaps in VTE risk knowledge; 2) risks and limitations of VTE prophylaxis and; 3) limitations of current VTE risk assessment models.[37, 40]

1.2.1.5 VTE Pathology and Complexity of Known Risk Factors

The pathology of HA-VTE is complex, involving interactions between clinical risk factors and acquired and/or inherited susceptibilities to thrombosis. Hemostasis is a normal physiologic

response to stop or prevent bleeding within a blood vessel. This innate response, mediated by the coagulation cascade, helps ensure blood vessel integrity and blood fluidity, however, hemostasis abnormalities can result in hemorrhage or thrombosis formation. While thrombi can form in both veins (VTE) and arteries (arterial thrombus), the two vascular disorders have different pathologies, clinical presentations, and outcomes.[41] VTE is associated with three conditions, known as Virchow's Triad, that predispose venous thrombus formation: hypercoagulability, venous stasis, and endothelial damage (Figure 3).

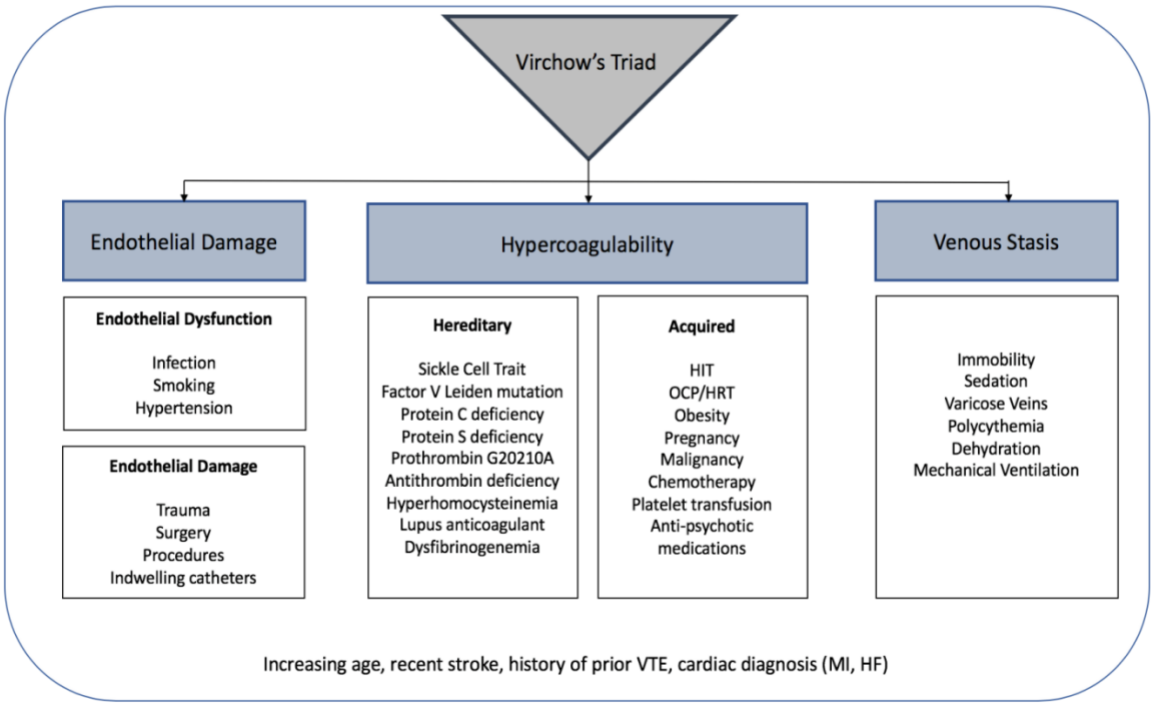


Figure 3. Major Risk Factors for VTE

The deep veins of the lower extremities are the most common site (96%) of DVT formation, specifically the veins of the thigh (iliac and femoral veins), the knee joint (popliteal vein), and the calf.[42] Within the calf, veins can be classified by those that receive blood flow from the foot sole (posterior tibial, anterior tibial, and peroneal veins) and the intramuscular veins (soleal and gastrocnemius veins).[42] With the assistance of valves, muscle compression of these deep veins

helps maintain the flow of venous blood in the lower extremities, and returns volume to the heart. Venous endothelium, as a result of avascular venous valves and deoxygenated venous blood flow, is predisposed to hypoxemia. Decreased mobility and/or poor baseline skeletal muscle function is further worsened by venous stasis resulting in the expression of adhesion molecules that increase activation of the coagulation cascade via the extrinsic pathway.

Iliac and femoral DVTs are commonly formed by iliac compression (such as a pelvic mass or trauma) and/or by catheter related vascular injury. As these vessels are large in diameter, thrombus occlusion often leads to obvious vascular insufficiency and clinical symptoms of pain, swelling, and erythema, and are associated with earlier clinician recognition and lower risk of embolizing to the lungs.[42] The veins of the calf, in contrast, have multiple branches and anastomoses. Thrombus formation in calf vessels are primarily due to lack of limb movement, are often without obvious clinical symptoms, and are more commonly associated with thrombus propagation and massive PE.[42] The vast majority of PE cases are secondary to an embolized DVT, however, case reports of de novo thrombosis of pulmonary arteries exist in trauma literature.[43-45].

1.2.1.6 VTE Risk Factors

Rates of VTE risk range from 1% in medical patients to as high as 40% in certain surgical and oncologic populations.[36, 46] Some VTE risk factors are associated with an increased overall lifetime risk (older age, venous insufficiency, obesity, as well as other circumstances which provoke acute risk (malignancy, recent surgery, indwelling vascular catheters, dehydration).[9, 47-59] Multiple risk factors are hypothesized to synergistically increase patient risk, but the relative contribution of each factor to overall VTE risk is unclear. Although VTE risk factors have been identified, a third to half of VTE cases are classified as unprovoked, meaning they occur in the absence of an identifiable provoking risk factor.[60]

Five major factors are known to be associated with inherited thrombophilia (factor V Leiden, prothrombin G20210A, protein C, protein S, and antithrombin levels),[61] although levels are not routinely evaluated in the absence of symptoms. Over the past two decades, genome-wide association studies have identified an expanding number of independent genetic variants associated with VTE. [62, 63] These studies have underscored the importance of inflammation in the pathophysiology of VTE, raised questions about the influence of heritability on VTE genotyped variants, and has identified some causal associations between blood traits and VTE.[61] While testing of single nucleotide polymorphisms associated with VTE is emerging, they are not yet relevant for meaningful use in clinical practice.

Symptoms of new-onset VTE in hospitalized patients often occur gradually over a period of hours to days and clinicians can easily fail to notice subtle feature pattern emergence until after a critical event occurs.[64, 65] As such, rates of HA-VTE are highly varied in studies where VTE diagnosis was made based on symptom assessment versus routine screening protocols.[66] Individual symptoms and risk factors for VTE have low predictive value (about 15%),[67] underscoring the need to identify dynamic composite feature pattern evolution associated with HA-VTE.

1.2.1.7 VTE Prophylaxis: Risks and Limitations

VTE prevention guidelines vary across patient populations.[68-70] Routine prophylaxis of VTE with low molecular weight heparin (LMWH) is a standard recommendation for hospitalized patients, however, it is not a panacea for prevention. Anticoagulant prophylaxis is associated with impaired wound healing, increased bleeding, need for blood transfusions and increased patient discomfort and costs, and can needlessly impose increased morbidity when applied indiscriminately to patients not at risk.[46, 71]

Conversely, underutilization of VTE prophylaxis has been linked to provider fear of these harms.[68-73] Identifying patient feature patterns predictive of VTE risk can inform more precise risk stratification and targeted prophylaxis application, helping those patients who need it most, but limiting needless exposure and iatrogenic complications in those who do not. *(Note: The candidate conducted a podium presentation on this topic at the American Association of Critical Care Nurses National Teaching Institute, 2019)*

1.2.1.8 Limitations of Current VTE Risk Assessment Models (RAM)

There is a lack of consensus regarding VTE clinical decision support tools.[68, 74] Published VTE RAMs, such as Caprini, Padua, and Wells are limited by their complexity of factors and lack of generalizability to different populations.[10, 11, 74] Some models include specialized coagulopathy or serum studies (i.e., lupus anticoagulant, prothrombin 20210A, homocysteine) not routinely ordered on all patients. After hospital admission, patients accrue conditions and undergo interventions which can alter baseline admission and lifetime risk predispositions to thrombosis.[37] All current RAMs share a common and significant limitation: they rely primarily on static, baseline patient features to assess risk for VTE. Hospital acquired VTE risk is associated with interactions between established and acutely acquired risk factors and pathology and symptoms evolve over the course of hospitalization. However, no current RAMs incorporate the progressive accrual of dynamic patient data and pattern evolution over hospitalization in their modeling approaches. *(Note: the candidate conducted a podium presentation on this topic at the Pennsylvania Coalition of Nurse Practitioners Annual Conference, 2018).*

1.2.1.9 Addressing the Gap with Machine Learning Approaches

VTE risk changes over the time-course of hospitalization and closing this gap in risk-knowledge and quantification requires the exploration of dynamically accruing clinical data. The totality of routinely collected EHR data over hospitalization duration holds potential for informing this gap, and data driven machine learning (ML) is the preferred method to develop and deploy predictive analytics for such data. Routinely used in other industries critically dependent on real-time analytics of massive amounts of disparate data (e.g., financial and retail), ML application in health care is nascent. ML methods are particularly suited for forecasting outcomes that change over time or vary from one case to another, as is common with hospitalized patients.[75] ML methodology has the ability to scale up correlational analyses to highly multivariate, high-frequency data to discern emerging complex patterns and relationships associated with disease evolution or clinical deterioration.[76] ML holds the potential to 1) identify complex mechanisms underlying disease; 2) lead to the generation of new hypotheses for research; and 3) inform improvements to existing explanatory models.[77, 78] ML approaches have been used to successfully develop highly sensitive models to predict risk of cardiorespiratory instability, cardiac ischemia and arrhythmia, and inform decision making to improve care delivery and patient outcomes.[75, 79-83] ML approaches can permit better understanding of the complex mechanisms underpinning new-onset, HA-VTE and improve current RAMs by incorporating dynamically accruing clinical data; improvements that will exert a sustained influence in this field of science.

1.2.1.10 The Value of Computable Phenotype Definitions

Research using health data originally obtained for clinical purposes requires accurate identification of specific health outcome cases and controls [84] from EHR or clinical database data. Diagnosis codes (assigned using the International Classification of Diseases, Ninth and Tenth

Revision, Clinical Modification [ICD-9-CM/ICD-10-CM] system) can be biased and inaccurate and they lack precise diagnosis times needed for rigorous approaches to predictive modeling. New-onset, HA-VTE has been identified as a challenging condition to identify due to its insidious and temporally evolving nature.[85] Computable phenotypes, computerized definitions of clinical conditions using standard data elements commonly available in EHR systems, offer a more rigorous approach to identifying patient records in big data repositories.[85-87] Although computable phenotypes can support cost-efficient and time-efficient reproducible queries,[88, 89] developing these tools is a significant informatics challenge. Because of the heterogeneity and dynamic nature of clinical data, the process of first identifying and defining a phenotype is a labor intensive, highly manual process conducted by domain experts. Once identified, translating human-readable phenotype information into a computable format that can be used across data sets and institutions requires a multi-disciplinary team that includes domain experts, biostatisticians, EHR informaticians, NLP experts, and computer scientists working in close collaboration.[90, 91] Although development of a computable phenotype for HA-VTE exceeds the scope of this dissertation study, the manual process of ground truth case ascertainment of new onset, HA-VTE will identify data elements associated with this condition that will contribute valuable information toward defining this phenotype in clinical data.

1.2.2 Innovation

This study will be the first to incorporate the progressive accrual of multi-domain, multi-granular intensively collected time series data in the development of ML models to predict dynamically evolving new-onset VTE risk in hospitalized patients. Identifying complex feature patterns predictive of new-onset HA-VTE risk will inform critical gaps in our understanding of

this common disease process and inform future study in which clinicians use this information to inform targeted prevention and treatment decisions. Finally, this study will be the first to exploit ML approaches to inform the critical practice of nursing surveillance for a nurse-sensitive patient complication.

1.3 Study Design and Methodology

1.3.1.1 Study Overview

The study design is an expanded and augmented secondary analysis of data originally obtained by Drs. Marilyn Hravnak and Michael Pinsky and their team of co-investigators Dr. Gilles Clermont, Dr. Artur Dubrawski and Ms. Melissa Saul: Predicting Patient Instability Non-invasively for Nursing Care (PPINNC). The IRB approved parent study, PPINNC (R01 NR01391), utilized machine learning analytics to develop models predictive of cardiorespiratory instability events in stepdown unit (SDU) patients. PPINNC data includes high frequency vital sign (VS) data which is rare outside of the intensive care unit and needed for complexity model building. The proposed research is sufficiently distinct from the parent project in that it will use the PPINNC data to answer a different research question and to achieve a different set of aims. Extraction of additional EHR data elements to augment the sample, annotation specific to VTE, and curation of a new-onset HA-VTE data set to serve as the machine learning platform, will be required to achieve these aims.

1.3.1.2 Construction of the Source Population

Time stamped VS data containing heart rate (HR), respiratory rate (RR), peripheral capillary oxygen saturation (SpO_2) and blood pressure (BP) data collected at 20 second intervals from Philips bedside monitors were originally collected for the parent study (total of 172,000 monitoring hours). Artifact detection algorithms were developed in the parent study to accurately identify artifactual vs. normal VS exceedances (excursion beyond a normal range),[13] providing a high degree of confidence that risk models are developed from artifact-free VS. Additional data were extracted to the study server from the UPMC Medical Archival Retrieval System (MARS).[92] These data include clinical data (time-stamped lab results, medications, microbiology culture results) and demographic and administrative data (age, gender, race, ICD-9-CM diagnosis and procedure codes, charge transactions) associated with each patient visit. All data were de-identified for the parent study.

1.3.2 Preliminary Work

It was recognized early on in the development of this research proposal that the cleaning, annotation and organization of data required to achieve Aim 1 rendered it the most challenging and time intensive aim and that the completion of Aims 2 and 3 relied on its achievement. Machine learning models that provide robust and actionable insights are predicated on data veracity and an established predictive relationship between some variables of interest.[77, 78, 93] Thus, exploring variable relationships as well as establishing prevalence of HA-VTE in the sample data set was essential information to determine feasibility. Preliminary work completed to establish these fundamental premises was completed to support the successful submission of an F31 NINR

training grant proposal and the findings from these preliminary studies also contribute to the completion and results of the first two aims.

1.3.2.1 Preliminary Study 1: Identification of a Study Population for New-Onset, HA-VTE Including Case Ascertainment[1]

A retrospective analysis of EHR data from 3680 SDU patients ages ≥ 22 years, was conducted to identify HA-VTE (either DVT or PE) cases. Although administrative coding, such as ICD-9-CM/ICD-10) is often used to identify disease conditions for billing purposes, its use in accurately identifying VTE cases and controls in clinical data, is unreliable.[94, 95] We therefore determined to identify new-onset, HA-VTE ground truth cases and controls using a multi-pronged manual review process of not only administrative coding[96] (Table 1), but also radiologic reports, and discharge summary reports. Admission and discharge diagnosis codes were reviewed for VTE identification. Unstructured (free text) radiologic reports for gold standard VTE diagnostic tests,[97] lower extremity Doppler ultrasound (LEDUS), computed tomographic angiography (CTA), ventilation-perfusion scan (VQ) and/or magnetic resonance angiography (MRA) performed at any point during the participants' hospital stay (n=4544 reports) were extracted from the MARS data repository. These reports were reviewed to determine ground truth new-onset HA-VTE cases. Our review protocol included preliminary screening of the diagnostic test reports by a bioinformatics

1 Portions of this preliminary study have been previously published as Pellathy, T., Saul, M., Clermont, G., Nagpal, C., Dubrawski, A., Pinsky, M., & Hravnak, M. (2018). 205: Accuracy of Identifying Venous Thromboembolism by Administrative Coding Compared to Manual Review. Critical Care Medicine, 46(1), 85. DOI: 10.1097/01.ccm.0000528224.97123.e0

expert using terminology extraction followed by manual expert review to identify positive new-onset VTE diagnostic test results and negative VTE test results occurring during the SDU LOS. Participants with chronic VTE or new-onset VTE occurring prior to SDU admission were identified and excluded. For indeterminate cases, we also extracted daily progress notes from MARS and reviewed them manually for language which could clarify the result. This rigorous and reproducible review process established ground truth outcome cases and quantified the prevalence of new-onset VTE in the overall sample (1.6%), and among patients with VTE diagnostic tests (10%).[98] Such prevalence is adequate for ML modeling.

Table 1. ICD-9 Codes Associated with Venous Thromboembolism (VTE)

ICD-9 code	Diagnosis
	Acute VTE
PE	
415.1	Pulmonary embolism and infarct
415.11	Iatrogenic pulmonary embolism and infarct
415.13	Saddle embolus
415.19	Other pulmonary embolism and infarct
DVT	
451.1 & .11	Phlebitis & thrombophlebitis of deep veins of lower extremity/femoral vein
451.19	Phlebitis & thrombophlebitis of deep veins of lower extremity, other
451.2	Phlebitis & thrombophlebitis of deep veins of lower extremity, selected.
451.81	Phlebitis & thrombophlebitis of other sites, iliac vein
451.89	Phlebitis & thrombophlebitis of other sites
451.9	Phlebitis & thrombophlebitis of unspecified site deep vein thrombosis, not of upper extremity
453.2	Other venous embolism and thromboembolism of vena cava
453.4-42	Acute venous embolism & thrombosis of deep vessels of lower extremity
453.8-453.89	Acute venous embolism and thrombosis of other specified veins
453.9	Other venous embolism & thrombosis of specified veins
	Chronic VTE
PE	
416.2	Chronic pulmonary embolism
DVT	
453.3	Chronic venous embolism and thrombosis of deep vessels of lower extremity
453.5	Chronic venous embolism and thrombosis of unspecified deep vessels of lower extremity
453.51	Chronic venous embolism and thrombosis of deep vessels of proximal lower extremity
453.52	Chronic venous embolism and thrombosis of deep vessels of distal lower extremity
453.7	Chronic venous embolism and thrombosis of other specified vessels
453.71	Chronic venous embolism and thrombosis of superficial veins of upper extremity
453.72	Chronic venous embolism and thrombosis of deep veins of upper extremity
453.73	Chronic venous embolism and thrombosis of upper extremity, unspecified
453.74	Chronic venous embolism and thrombosis of axillary veins
453.75	Chronic venous embolism and thrombosis of subclavian veins
453.76	Chronic venous embolism and thrombosis of internal jugular veins
453.77	Chronic venous embolism and thrombosis of other thoracic veins
453.79	Chronic venous embolism and thrombosis of other specified veins

This work was presented nationally in poster format at the Society for Critical Care Medicine (SCCM) Annual Congress 2018, the abstract published, and a data-based manuscript (Appendix B: Dissertation Manuscript 2) containing data from this preliminary study, as well as the final results from study Aim 2 is currently under review.

1.3.2.2 Preliminary Study 2: New-Onset, HA-VTE Model Development [2]

Gold standard test confirmed positive and negative new-onset VTE cases were linked with demographic and mean VS data (HR, RR, BP, SpO_2) 48 hours preceding diagnosis, producing a final sample of 552 unique patients. A logistic regression was fit to evaluate the relationship of mean VS data 48 hours preceding VTE diagnostic testing to the likelihood of new-onset VTE diagnosis. The predictive model was statistically significant ($\chi 2(6) = 5.875, p = .003$), with an area under the receiver operating curve (AUC) .594, 95% CI (.511-.696) demonstrating increasing mean HR is significantly associated with an increased likelihood of a new-onset, HA-VTE diagnosis (Table 2).[98] These preliminary analyses, conducted with a small quantity (48 hours) of mean VS data identified that even coarse features (summary statistics of longitudinal data) were predictive of new-onset, HA-VTE, and demonstrated beginning skill in ML approaches. Based on initial proof that even mean VS data can predict the VTE outcome, the proposed study will further predictive model development on multidimensional, intensively collected time series data with a greater degree of granularity and using a variety of ML approaches.

[2] Portions of this preliminary study have been previously published as Pellathy, T., Chen, L., Dubrawski, A., Clermont, G., Pinsky, M., & Hravnak, M. (2018, March). Prevalence of Venous Thromboembolism (VTE) in an Adult Step-Down Unit Population: A Proof-of-Concept Feasibility Study for Machine Learning Predictive Model Development. Nursing Research (Vol. 67, No. 2, pp. E130-E130). Two Commerce Sq, 2001 Market St, Philadelphia, Pa 19103 USA: Lippincott Williams & Wilkins.

Table 2. The Predictive Relationship Between Patient Vital Signs and VTE Risk in Step Down Unit Patients

	B	SE	Wald	*df*	*p*	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Respiratory rate	.061	.044	1.943	1	.160	1.064	.976	1.160
Heart rate	**.024**	**.010**	**5.808**	**1**	**.016***	**1.025**	**1.005**	**1.045**
SpO2	-.034	.071	.227	1	.634	.967	.842	1.111
Systolic BP	.005	.012	.200	1	.655	1.005	.983	1.028
Diastolic BP	.005	.019	.059	1	.808	1.005	.968	1.043
Constant	-4.86	.945	26.48	1	.000	.008		

*Significant at $p < 0.05$

This work was presented in poster format at the Eastern Nursing Research Society Annual Meeting, 2018, and the abstract published.

1.3.2.3 Preliminary Study 3: Identifying Time-Interval Features Predictive of HA-VTE[3]

We know ML approaches can identify patterns predictive of disease in time series multidimensional clinical data that cannot be elicited with traditional statistical approaches. As HA-VTE risk factors and disease pathology are dynamic over time course of hospitalization, we sought to develop a preliminary model incorporating the progressive accrual of patient data over the time course of hospitalization in the modeling approach. We conducted a retrospective analysis of 3680 adult SDU patients to identify confirmed VTE positive (n=53) and negative (n-592) cases. These cases were linked with medication and laboratory counts (MEDLAB) preceding diagnosis (positive/negative). Medication data was based on drug category and number of doses and lab data was based on frequency of lab test and normal/abnormal results of those tests. MEDLAB data for

3 Portions of this preliminary study have been previously published as Pellathy, T., Chen, L., Clermont, G., Dubrawski, A., Pinsky, M., & Hravnak, M. (2019). 58: Identifying Time Interval Features Predictive of Hospital-Acquired Venous Thromboembolism. Critical Care Medicine, 47(1), 29. DOI: 10.1097/01.ccm.0000550850.41971.6a

48 hours preceding HA-VTE positive or negative diagnoses were used in all models.

These data were then combined with featurized VS data time windows of 12, 24, 48, and 72 hours and binary logistic regression was employed to evaluate the ability of MEDLAB and VS data at different time intervals to predict new onset, HA-VTE.[99]

The models using data 72 hours and 12 hours preceding new-onset, HA-VTE diagnosis were not significant, possibly due to the bluntness of the featurized variables and data sparsity. The model developed with data 48 hours (Figure 3, blue data) preceding new-onset, HA-VTE diagnosis had a sensitivity of 63% and specificity of 93%. Variables in the model that were significant ($p < 0.05$) included the VS variables of HR, RR, and systolic BP. Also significant were patients receiving selective serotonin reuptake inhibitors, antipsychotics and erythropoiesis-stimulating agent medications. Pro-thrombin time and pro-time counts, meaning how frequently these lab tests were ordered and resulted, were also significant in this model.

The model with vital sign data 24 hours preceding new-onset, HA-VTE diagnosis (Figure 4, red data) was slightly improved with regard to sensitivity. Sensitivity increased from 63% in the 48-hour model to 73%. Specificity remained the same as the 48-hour model (93%). Notably, variables in the 24-hour model that were significant ($p < 0.05$) varied slightly from the 48 model. Respiratory rate and blood pressure remained significant; however, heart rate was not a significant predictor 24 hours preceding VTE diagnosis. The medication and lab variables significant in the 48- hour model, were also significant in the 24-hour model. Also significant were steroid medications and laboratory band count.

These findings demonstrate patient features that significantly predicted new-onset, HA-VTE can change over a time. Furthermore, these findings demonstrate a significant relationship

between variables that supports the value of continued work utilizing machine learning algorithms for predictive modeling of new-onset, HA-VTE.

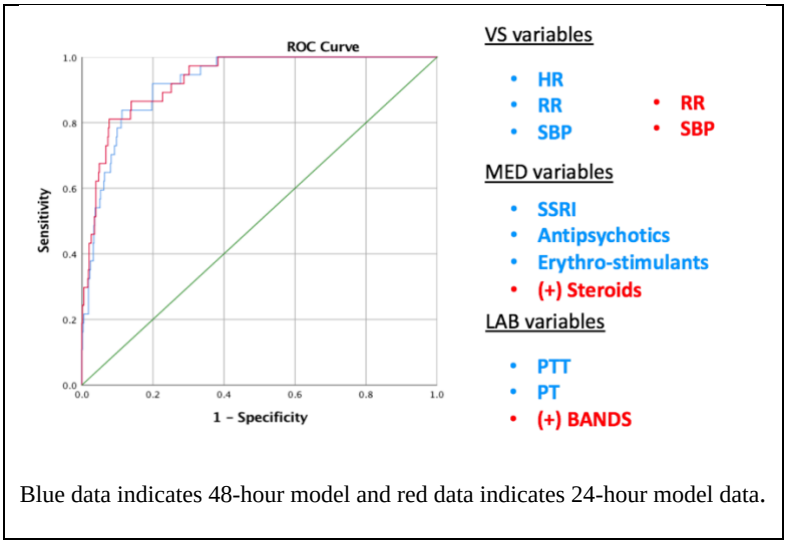


Figure 4. Results of MEDLAB + 48-hour and 24-hour Vital Sign Windows.

This work was presented as a podium presentation at the Society of Critical Care Medicine Annual Congress 2019, received a STAR research award, and the abstract published.

1.4 Setting and Sample

1.4.1 Sample Entry/Exclusion Criteria

Entry criteria for the parent study were patient need for a monitored bed in the SDU of the University of Pittsburgh Medical Center (UPMC) and age >21 years between 11/06 and 09/08. Under Institutional Review Board (IRB) approval for waiver for informed consent, every patient admitted to the 22-bed medical-surgical-trauma SDU during the study timeframe contributed to the data without exclusion (except age <21). No special classes of patients were excluded. These

practices yielded a convenience sample of 3680 patients. Patients remained in the study until discharge to another unit or from the hospital. Data collection for each patient spanned the entire duration of the patient's SDU length of stay (LOS), for the total unit census across the study time frame. Patients discharged or transferred to a non-study unit were discontinued from the study, but their data prior to that time retained. The proposed study population is obtained from this source population.

The proposed study aims to develop models predictive of new-onset, HA-VTE, defined as VTE verified by a gold-standard diagnostic test between the time of SDU admission and time of SDU discharge. Patients with a diagnosis of chronic VTE, determined either by admission diagnosis or radiology report findings, will be excluded. Patients with new-onset, HA-VTE identified during the hospitalization but before SDU admission or >24 after SDU discharge will also be excluded, as the continuous VS data streams necessary for the modeling are not available for those patients.

During the study time frame, some participants were admitted to the SDU more than once during separate hospital admissions and/or during the same hospital LOS. For this analysis, each SDU admission LOS (and all clinical and VS data streams during that admission) is considered a case. This means some study participants will have contributed to more than one case used in analysis. As we are looking at clinical patient features over time of SDU admission that are predictive of new-onset, HA-VTE, this approach is consistent with our research question and aims and does not result in any overlapping data use.

1.4.2 Characteristics of the Study Sample

The overall patient sample is consistent with the characteristics of adult patients admitted to the study unit during the study period (75% white, 71% male, mean age 58.0 years). Further sample demographics are detailed in the analysis and results sections that follow.

1.4.3 Sample Size Justification

1.4.3.1 ML Sample Size

ML methods do not require a justification of sample size. Of greatest importance is the prevalence of the outcome variable of interest in the sample. Validity and prevalence of the outcome variable (new-onset VTE) has been determined in preliminary work (1.3.2.1), are consistent with VTE prevalence rates reported in the literature,[7] and are adequate for ML accuracy. [100, 101]

1.4.3.2 Binary Logistic Regression Sample Size

Prior to applying ML algorithms, binary logistic regression (BLR) will be used to explore underlying assumptions between patient factors and VTE diagnosis. Odds ratios for VTE risk in the literature vary from 1.34 in extremely low risk patients (ambulatory, community dwelling) to as high as 9.06 in very high-risk patients (septic, orthopedic, oncology)[102, 103]. To be conservative, an odds ratio of 2.00, which translates to a medium effect size, was chosen for sample size determination.

Given this is a retrospective study and sample sizes are established based on data availability, a post-hoc power analysis was conducted for analysis that employs logistic regression

of a binary dependent variable for model building. For a two-tailed test with alpha set at 0.05, power analysis was conducted to assess power based on a medium and small effect size for our sample sizes of 630 and 2110 that will be detailed in a subsequent section (1.5.3). Power exceeds 80% for both samples and effect sizes.[104]

1.4.4 Variables

1.4.4.1 Dependent Variable (DV)

The DV in modeling is new-onset, HA-VTE diagnosis and its onset time, defined as the time the new-onset VTE diagnostic confirmatory test was conducted. (Table 3)

1.4.4.2 Independent Variables (IV)

The IV's are patient data associated with new-onset, HA-VTE signs, symptoms or risk factors identified in the literature, that can be further categorized as dynamic or static. Dynamic variables are patient data marked by activity or change during the hospital admission up to the time of VTE diagnosis will include: continuous VS of HR, RR, BP and SpO_2 measured in 20 second intervals, laboratory values, medication categories, blood and blood product transfusions, procedures, and radiology and ultrasound tests (all time-stamped). Static variables are those data determined upon hospital admission and include: ICD-9-CM admission diagnosis, age, gender, race, height and weight. A listing of all variables and their proposed featurization can be found in Table 4. A breakdown of the individual medications in each drug category can be found in Appendix C. These variables are informed by existing VTE RAMs and published research on factors associated with VTE risk. Clinical and demographic variables available in this sample were

mapped to currently used RAMs and details can be found in Appendix D. The continuous VS data were collected in the parent study, using Philips bedside monitor (Andover, MA).

Table 3. Dependent Variables

Variable	Level of Measurement	Data Source	Measures	Descriptive Statistics
New-onset VTE	nominal	MARS	Tested, confirmed positive; Tested, confirmed negative; Never tested	Frequencies, percentage, mode
Time of VTE diagnosis	ratio	MARS	Time from SDU admission to confirmatory diagnostic test, in minutes	Mean, standard deviation (SD), range, minimum (Min), maximum (Max), skewness, kurtosis

Table 4. Independent Variables

Variable	Level of Measurement	Data Source	Measures	Descriptive Statistics
Demographics				
Age	ratio	MARS	Calculated from patient date of birth to admission	Mean, standard deviation (SD), range, minimum (Min), maximum (Max), skewness, kurtosis
Gender	nominal	MARS	Registration data	Frequencies, percentile, mode
Race	nominal	MARS	Registration data	Frequencies, percentile, mode
Charlson Comorbidity Index Deyo Method (CCI-D)	interval	MARS	Score calculated from ICD-9-CM hospital discharge diagnoses	Mean, SD, range, Min, Max, skewness, kurtosis

Step-Down Unit length of stay (LOS)	ratio	MARS	Minutes (converted to days and hours as appropriate)	Mean, SD, range, Min, Max, skewness, kurtosis
Inpatient LOS	ratio	MARS	Days	Mean, SD, range, Min, Max, skewness, kurtosis
Vital Signs				
Heart rate (HR)	ratio	Philips bedside monitor	Beats/minute	Mean, SD, range, Min, Max slope, skewness, kurtosis
Respiratory rate (RR)	ratio	Philips bedside monitor	Breaths/minute	Mean, SD, range, Min, Max slope, skewness, kurtosis
Systolic Blood Pressure (SBP)	ratio	Philips bedside monitor	Millimeters of mercury (mmHg)	Mean, SD, range, Min, Max slope, skewness, kurtosis
Diastolic Blood Pressure (DBP)	ratio	Philips bedside monitor	mmHg	Mean, SD, range, Min, Max slope, skewness, kurtosis
Mean Arterial Pressure (MAP)	ratio	Philips bedside monitor	mmHg	Mean, SD, range, Min, Max slope, skewness, kurtosis
Pulse Oximetry (SpO_2)	ratio	Philips bedside monitor	Percent oxygen saturation	Mean, SD, range, Min, Max slope, skewness, kurtosis
Diagnoses				
Myocardial Infarction (MI)	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Congestive Heart Failure (CHF)	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Stroke	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Malignancy	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Lower Extremity Fracture	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Diabetes Mellitus	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Obesity	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Hypertension (HTN)	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Chronic obstructive pulmonary disease (COPD)	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Malignancy	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Thrombophilia	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Septicemia	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Chronic Venous Insufficiency	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Dehydration	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode

Chronic kidney disease	nominal	MARS	ICD-9-CM code	Frequency, percentile, mode
Medication Classes				
Anti-coagulant	ratio	MARS	Charge count/24 hrs*	Mean, SD, range, Min, Max
Anti-infective	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Anti-platelet	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Anti-psychotic	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Benzodiazepine	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Cardiac	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Erythro-stimulant	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Gastrointestinal	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Hormone	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Hyperglycemic	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Immunosuppressant	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Naloxone	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Narcotic	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Neuro	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Non-steroidal Anti-inflammatory (NSAID)	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Sedative	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Selective Serotonin Reuptake Inhibitors (SSRI)	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Steroids	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Thrombolytic	ratio	MARS	Charge count/24 hrs	Mean, SD, range, Min, Max
Laboratory Values				
Pro-thrombin (APT) time, mean	ratio	MARS	Seconds	Mean, SD, range, Min, Max
APT normal/ abnormal high/ abnormal low count	ratio	MARS	Test count**	Mean, SD, range, Min, Max
Partial thromboplastin time (PTT), mean	ratio	MARS	Seconds	Mean, SD, range, Min, Max
PTT normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Bands, mean	ratio	MARS	x 10^3 /mm^3	Mean, SD, range, Min, Max

Bands normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Brain natriuretic peptide (BNP), mean	ratio	MARS	picograms per milliliter (pg/ml)	Mean, SD, range, Min, Max
BNP normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Cancer antigen 125 (CA-125), mean	ratio	MARS	units/mL	Mean, SD, range, Min, Max
CA-125 normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Carcinoembryonic antigen (CEA), mean	ratio	MARS	ng/ml	Mean, SD, range, Min, Max
CEA normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Serum Carbon Dioxide (CO_2), mean	ratio	MARS	mmol/L	Mean, SD, range, Min, Max
CO_2 normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Creatinine, mean	ratio	MARS	mg/dL	Mean, SD, range, Min, Max
Creatinine normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
C-reactive Protein (CRP), mean	ratio	MARS	mg/L	Mean, SD, range, Min, Max
CRP normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Direct Bilirubin (DBILI), mean	ratio	MARS	mg/dL	Mean, SD, range, Min, Max
DBILI normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Sedimentation Rate (ESR)	ratio	MARS	mm/h	Mean, SD, range, Min, Max
ESR normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Iron (Fe), mean	ratio	MARS	mcg/dL	Mean, SD, range, Min, Max

Fe, normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Ferritin, mean	ratio	MARS	ng/mL	Mean, SD, range, Min, Max
Ferritin normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Glycosylated Hemoglobin (HgA1c)	ratio	MARS	%	Mean, SD, range, Min, Max
HgA1c normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Hematocrit (HCT), mean	ratio	MARS	%	Mean, SD, range, Min, Max
HCT normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Hemoglobin (Hgb), mean	ratio	MARS	g/dL	Mean, SD, range, Min, Max
Hgb normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Potassium (K), mean	ratio	MARS	Mmol/L	Mean, SD, range, Min, Max
K, normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Lactate, mean	ratio	MARS	IU/L	Mean, SD, range, Min, Max
Lactate, normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Lipase, mean	ratio	MARS	U/L	Mean, SD, range, Min, Max
Lipase, normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Magnesium (Mg), mean	ratio	MARS	mmol/L	Mean, SD, range, Min, Max
Mg normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
$PaCO_2$	ratio	MARS	mm/Hg	Mean, SD, range, Min, Max
$PaCO_2$ normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Platelet (PLT) count, mean	ratio	MARS	cells/*u*L	Mean, SD, range, Min, Max
PLT normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
PaO_2	ratio	MARS	mm/Hg	Mean, SD, range, Min, Max
normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max

Total Thyroxine (T4)	ratio	MARS	mcg/dL	Mean, SD, range, Min, Max
T4 normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Triglyceride (TRG), mean	ratio	MARS	mg/dL	Mean, SD, range, Min, Max
TRG normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Troponin, mean	ratio	MARS	ng/mL	Mean, SD, range, Min, Max
Troponin, normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Thyroid Stimulating hormone (TSH), mean	ratio	MARS	mU/L	Mean, SD, range, Min, Max
TSH, normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Total triiodothyronine (T3), mean	Ratio	MARS	ng/dL	Mean, SD, range, Min, Max
T3, normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
White Blood Cell (WBC), count	ratio	MARS	x 10^9/L	Mean, SD, range, Min, Max
WBC normal/ abnormal high/ abnormal low count	ratio	MARS	Test count	Mean, SD, range, Min, Max
Procedures				
Central Venous Catheter	nominal	MARS	Inserted yes/no	Frequencies, percentage, mode
Surgical status	nominal	MARS	Non, major, minor	Frequencies, percentage, mode
Type of Surgery	nominal	MARS	GI, CABG, Vascular, Orthopedic, Trauma, Neuro, Thoracic, Gynecologic malignancy, Other	Frequencies, percentage, mode

MARS Medical Archival System, $PaCO_2$ partial pressure of arterial carbon dioxide, PaO_2 partial pressure of arterial oxygen, CABG coronary artery bypass graft

*Charge count = Count of medication charge transaction doses ordered/patient/24 hours

**Test count = Count of tests with normal/ abnormal low/ abnormal high results

1.5 Methods

Preliminary and proof-of-concept work (Section 1.3.2) identified a gold-standard study population for new-onset, HA-VTE cases and established scientific merit and feasibility for the now proposed study described below. Developing models to identify complex patterns of patient features that can inform evolving VTE risk requires transformation of a great quantity of disparate, intensely longitudinal EHR data into a data platform suitable for ML analysis and model building. This will be achieved through the following aims (Figure 5):

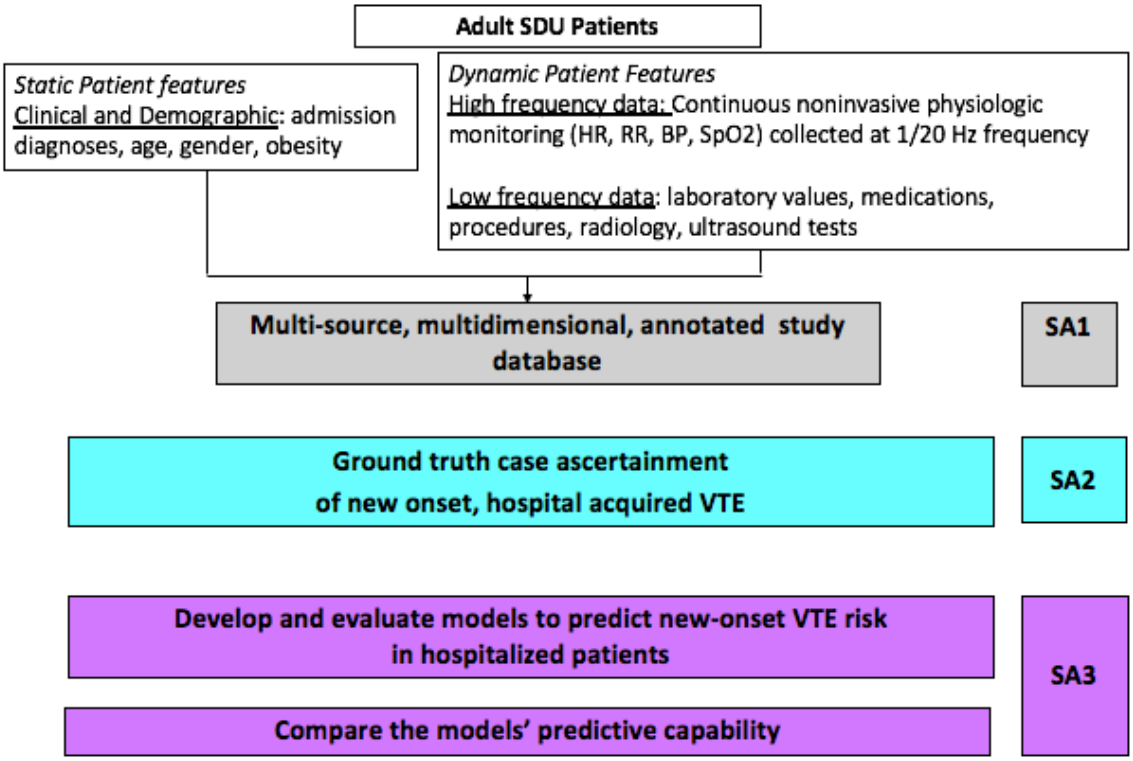


Figure 5. Study Model

1.5.1 Specific Aim 1: Assemble a Large Scale, Multi-Source, Multidimensional Dataset from an Adult Hospital Step Down Unit (SDU) Sample.

A large scale, multi-source, multidimensional dataset will be constructed to serve as the new-onset, HA-VTE ML platform. Beyond available data from the parent study, the applicant will extract additional structured and unstructured EHR data (radiology, ultrasound and transfusion data) and annotate previously extracted structured data (medications, laboratory data, procedure data), to augment the static and dynamic patient data points from the parent study (1.3.1.2). For each participant, we will then link the static and time-stamped dynamic patient data from the sources outlined, producing an intensive longitudinal dataset affording the ability to examine variables at different time windows prior to new-onset, HA-VTE occurrence (VTE positive diagnostic test time stamp).

1.5.1.1 Data Annotation and Curation.

EHR data streams must undergo preparation for computational model building. In this step, extracted data will be transformed into an annotated clinical research data set. To ensure study rigor and optimize reproducibility, data elements described will be cleaned, annotated and reported in accordance with the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) international guidelines for studies conducted using routinely-collected health data [105, 106]. As indicated, standard annotation vocabularies will be used: Systematized Nomenclature of MEDicine - Clinical Terms (SNOMED CT) for free text radiology and clinical notes, RxNORM for medications, Logical Observation Identifiers Names and Codes (LOINC) for laboratory and radiology tests and Centers for Medicare and Medicaid Services (CMS) ICD-9-CM codes for procedures and clinical diagnoses.[107-109]

1.5.2 Specific Aim 2: Ground Truth Case Ascertainment

Identification of cases and controls will be an iterative process informed by ground truth identification of the condition of interest followed by identification and annotation of EHR data elements associated with the condition. The first step, ground truth new-onset, HA-VTE identification by gold standard review was completed in pilot work (Section 1.3.2.1).

1.5.2.1 Radiologic Data Annotation and Time Stamping for Cases

Following the identification and exclusion of patients with chronic VTE or new-onset VTE occurring prior to SDU admission, all remaining SDU cases were annotated for one of three possible outcome events: 1) new-onset VTE positive, 2) tested for new-onset VTE and found to be negative by gold standard diagnostic test, and 3) not tested for new-onset VTE during SDU stay. (Figure 6)

1.5.2.2 Annotation of New-Onset, HA-VTE Positive Cases

Data was annotated for new-onset, HA-VTE time of occurrence and subtype (DVT or PE) to allow for maximal insight into the data during analysis. A new-onset DVT case is defined as an acute venous thrombosis in at least one of the following deep veins on LEDUS: internal jugular, superior vena cava, inferior vena cava, iliac, femoral, popliteal and profunda femoris veins. Chronic thromboembolic disease and thrombi in the following vessels are excluded from the definition of new-onset DVT: Portal circulation veins, superficial veins, man-made venous conduits and arteries. New-onset PE is defined as presence of an acute occlusive or sub-occlusive clot in a main, lobar, segmental and/or sub-segmental pulmonary artery on CTA or MRA. CTA and MRA results indicating a filling defect in the aforementioned vessels and V/Q results scans

with intermediate or greater probability findings will require further review of medical notes to confirm/exclude new-onset, HA-VTE diagnosis. Time stamp for the occurrence of a new onset, HA-VTE outcome event is annotated as time zero (T_0), the date and time the confirmatory diagnostic test was conducted. Indeterminate radiology test results subsequently confirmed as a new-onset VTE event, based on review of additional medical notes, were time stamped at the time of the indeterminate test.

1.5.2.3 Annotation of VTE Negative by Confirmatory Test Cases

SDU patients with a gold standard radiology test negative for the presence of embolus/emboli in specified vasculature were annotated as confirmed VTE negative cases (1.3.1.2). Time stamp for the occurrence of a confirmed VTE negative outcome event T_0 is the date and time the diagnostic test that ruled-out HA-VTE diagnosis was conducted. Indeterminate radiology test results subsequently confirmed as negative for VTE, based on review of additional medical notes, were annotated for T_0 at the time of the indeterminate test.

1.5.2.4 Annotation of Not Tested Cases

SDU patients who did not undergo a VTE diagnostic test during their SDU stay or during the 24-hours immediately after SDU discharge, are annotated as a not tested case. As this is a cohort of untested patients, a hypothetical outcome event time T_0 must be defined for these cases. The T_0 for not tested patients will be defined by the following two-step process:

1. The distribution of time from SDU admission to gold standard VTE diagnostic test time (time-to-test) for all tested patients (1.5.2.1 and 1.5.2.2) will be determined.
2. Not tested cases will be matched to the mean time-to-test distribution for all tested patients and then a random sample from that distribution will be obtained. This

enables evaluation of both groups at similar time points in the SDU stay. The rationale for defining and annotating two different types of control groups is outlined in detail in Section 1.5.3 below.

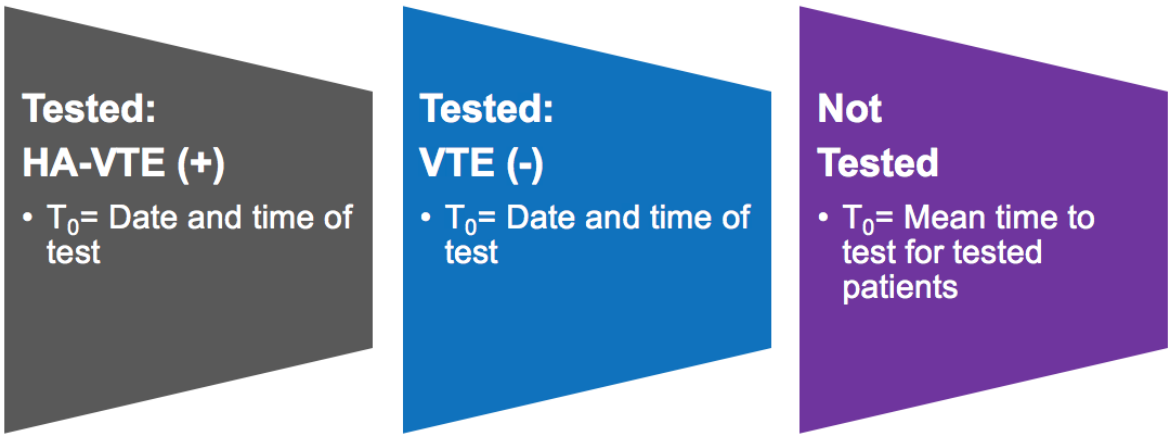


Figure 6. Each Case is Annotated for One of Three Outcome Events

1.5.3 Specific Aim 3: Develop and Evaluate Models to Predict New-Onset, HA-VTE Risk in Hospitalized Patients

Our approach to model building stems from the traditional ML processes of model training, validation and testing. In detail, we aim to capture all potentially informative patient variables and quantify their utility in predicting new-onset, HA-VTE diagnosis. Our approach has four stages: 1) Feature engineering, 2) Feature selection and model development, 3) Model performance evaluation, and 4) Model interpretation. Ground truth annotation and a discrete outcome allows us to formulate this as a supervised machine learning classification problem.

Aims 1(1.5.1) and 2 (1.5.2) comprise the stages of data processing required to produce a record of all such detected outcome events, annotated with time stamp of new-onset, HA-VTE

diagnosis, confirmatory VTE negative study, or absence of a VTE diagnostic test, event type descriptors (VTE subtype), time-stamped dynamic patient features (VS data and lower frequency clinical data) and static patient features (present on hospital admission data) throughout the entire SDU length of stay up to event occurrence, forming a transactional dataset.

1.5.3.1 Problem Formulation

Our problem formulation for ML model building uses the three case-cohorts defined in section 1.5.2 and illustrated in Figure 6 above, and is guided by the following clinical rationales:

1. Among all SDU patients, can we predict cases for which clinicians suspect new-onset HA-VTE pathology (and test) versus those they do not suspect (and not test)? This question aims to inform the challenge clinicians face regarding risk stratification of patients. Sample n = 2110, tested cases rate 30%, Figure 7.

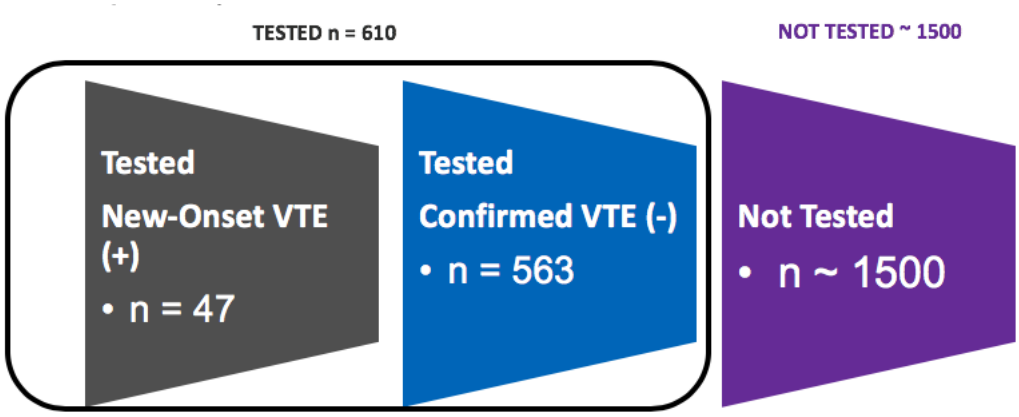


Figure 7. Case-Cohorts for Clinical Rationale #1

2. Among SDU patients who clinicians suspect new-onset, HA-VTE pathology, can we predict confirmed positive diagnosis from those whose diagnostic work up is confirmed to be negative? The ability to discern differences between 2 groups of hospitalized patients

who both carry high clinical suspicion for new-onset HA-VTE is an important clinical challenge. Sample n= 630, VTE (+) rate 8%, Figure 8.

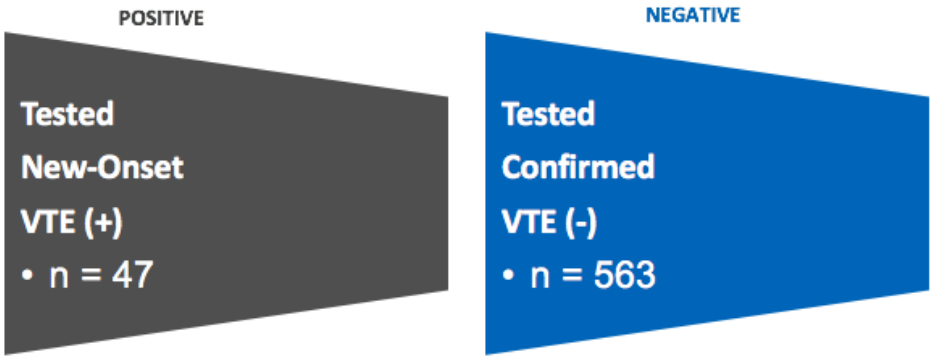


Figure 8. Case-Cohorts for Clinical Rationale #2

We hypothesize our models will be able to discriminate between positive and negative HA-VTE diagnosis and tested and untested patients in advance of T_0. Additionally, we hypothesize that the progressive addition of more granular time series data will both improve predictive ability as well as inform the relative importance of the features in the models.

1.5.3.2 Problem Formulation Constraints

Our approach to problem formulation and extraction of our time series features requires consideration of the following constraints:

New-onset VTE disease pathology. The exact time frame for hospital acquired VTE development is not precise. Hemostasis abnormalities leading to thrombus formation can occur gradually, with symptoms evolving over hours to days and the exact time frame for VTE development varies greatly by individual. [110]

Absence of published models forecasting HA-VTE risk. There is little published evidence to guide the optimal time intervals for forecasting HA-VTE risk. In the outpatient

setting, Posch et. al., recently explored the "dynamic" prognostic ability [111] of quantitative changes in D-dimer lab values to inform a patient's risk of cancer associated VTE and found monthly serum levels increased over a six-month time period in cancer patients who developed VTE and remained stable in cancer patients who did not develop VTE. In hospitalized patients, to the best of our knowledge, this is the first study to employ dynamic time series data gathered during hospitalization to predict risk of HA-VTE.

Absence of VTE risk assessment surveillance guidelines specific to the SDU population. There are no published time interval recommendations for new-onset VTE surveillance in the SDU population. Recommendations for the highest risk populations (neuro and multi-trauma ICU patients), include initial screening patients for the development of VTE within 48 hours of ICU admission followed by bi-weekly surveillance.[66, 112]

Data availability. Data to be employed in modeling is available in varying levels of granularity, which informs problem formulation. Static, present on admission data that is collected and available to clinicians at time of hospital admission includes demographic information and diagnosis information known at time of admission. Medication and procedure data, obtained from the charge data, are available with a time granularity of a calendar day. Laboratory data are available with a granularity of day and time resulted. Lastly, our dynamic VS data (HR, RR, SPO2) are available every 20 seconds and BP every 2-4 hours. The p4 phase has significant missing VS data due to a storage issue during data collection and this impacts data availability for this set of features for a number of patients in this study sample.

1.5.3.3 Problem Formulation Workflow

To test the hypothesis that the progressive addition of more granular time series data will improve model prediction performance, we will build our models in stages (Figure 9). In the first

stage, we propose to build a model with static, present on admission data only. In the second stage, we will add low frequency data and in the third and final stage, dynamic, higher frequency, VS data will be added.

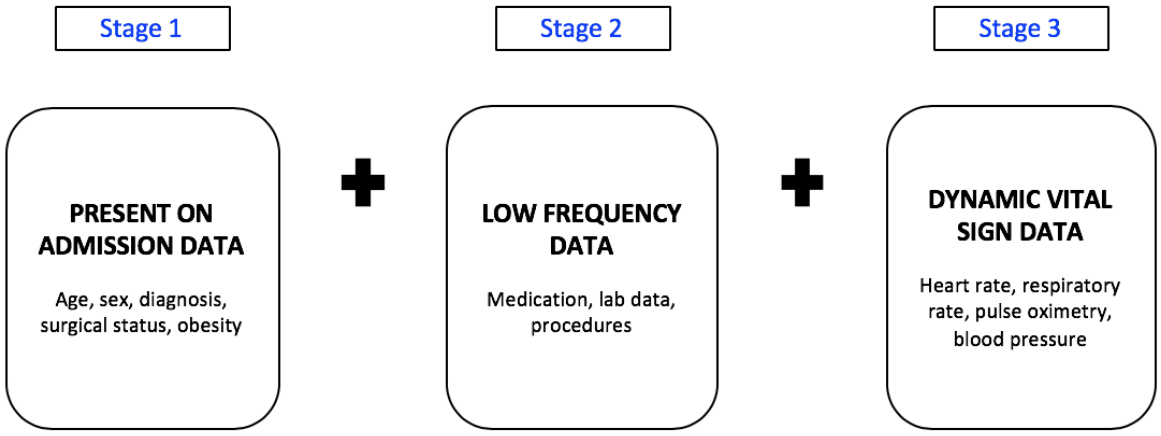


Figure 9. Model Building Stages

1.5.3.4 Machine Learning Process

As shown in Figure 10, our machine learning process design will involve periods of **observation** during SDU stay (during which potentially predictive VS and clinical events are extracted) and **prediction** (specified time windows preceding occurrence of HA-VTE). The models will incorporate dynamic and static data at various lead times (24 and 48-hours) before

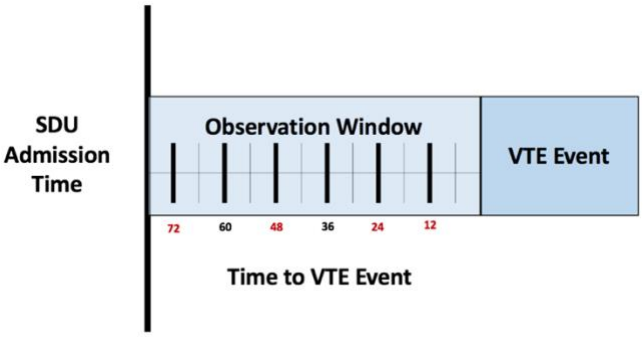


Figure 10. Machine Learning Process

the outcome VTE event. Starting with static data and adding increasingly dynamic time series data (Figure 9), one goal of our learning experiments will be to build the best attainable model for predicting the VTE outcome event at two times points: 1) time of diagnosis (T_0) and 2) 24-hours preceding diagnosis (T_{-24}). (Figure 11) These models will likely rely on different subsets of input features.

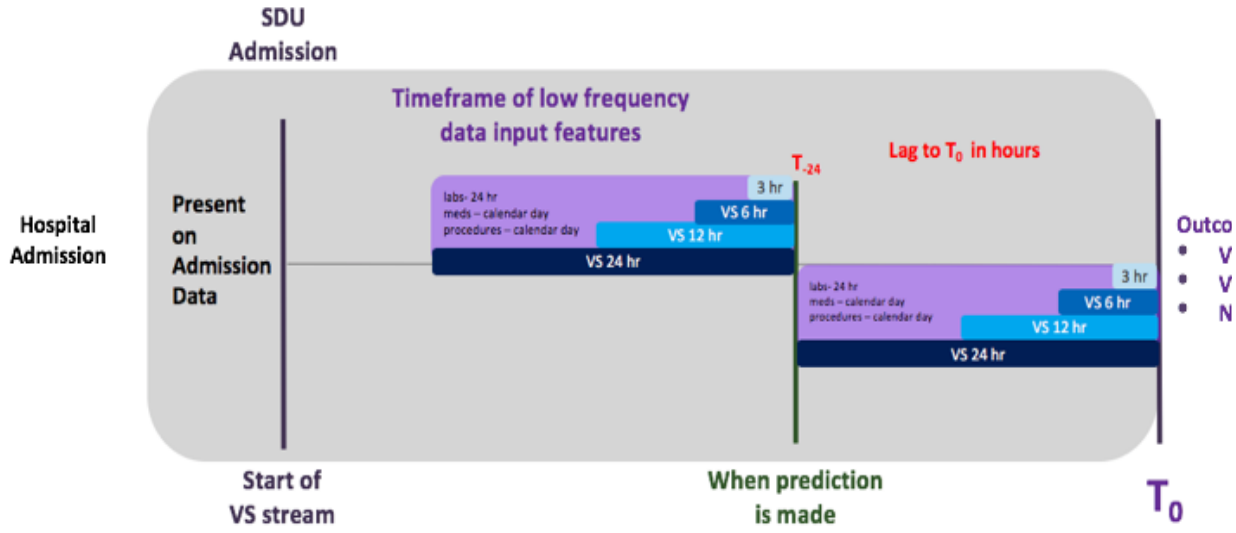


Figure 11. Machine Learning Process in Detail

Using the 3-stage approach outlined in Figures 9 and 11, for each clinical rationale and prediction time point (T_0 and T_{-24}), we will develop three sets of models.

Prediction time point T_0

1. **T_0 Static Models**: Static (present on admission) data
2. **T_0 Low Frequency Models**: Static + low frequency data
3. **T_0 High Frequency Models**: Static + low frequency + high frequency data

Prediction time point T_{-24}

1. **T_{-24} Static Models**: Static (present on admission) data
2. **T_{-24} Low Frequency Models**: Static + low frequency data
3. **T_{-24} High Frequency Models**: Static + low frequency + high frequency data

The proposed data features, candidate models, and feature selection approaches for model building are detailed in sections 1.5.4 and 1.5.5 below.

1.5.4 Specific Aim 3a*:* Apply ML Algorithms to Develop and Evaluate Models for New-Onset HA-VTE risk Over Time.

1.5.4.1 Feature Engineering

To maximize its potential, a ML model needs an information rich set of features. Domain knowledge and extensive review of the literature will inform feature engineering. Every effort will be made to ensure only variables available at the time of prediction are included and that these data are ones readily available in clinical practice. Furthermore, every effort will be made to limit the use of environmentally influenced variables. This data set will serve as the input to the second stage of processing: feature selection and model development. Proposed features for model building are detailed in Appendix E.

1.5.4.2 Candidate Model Selection

Candidate model selection is based on data characteristics and constraints (data missingness, level of measurement, class imbalance) and a desire to model interpretability. The bias-variance trade-offs of different models were also considered. The proposed candidate models and feature selection methods include:

- Binary logistic regression (BLR) is a simple classifier that can provide a good baseline model for comparison. Interpretability of BLR is an advantage, however this approach is likely to produce a model with higher bias (overfitting). LASSO (least absolute shrinkage and selection operator) will be used for feature selection.
- Naïve Bayes (NB) is another highly interpretable classifier that has been shown to perform well on small amounts of training data and to be relatively robust to missing data. Similar to BLR, NB tends to build models with higher bias, but low variance. As NB is a classifier that assumes all features are unrelated to each other (conditionally independent), we proposed to fit this model using all predictor variables without using feature selection for this modeling approach.
- Random Forest (RF) is an ensemble method that uses bagging (bootstrap aggregating) to combine many decision trees in parallel enabling its ability to produce models with low bias and moderate variance. A powerful classifier that is fairly robust to outliers and missing data, RF also handles categorical and continuous data well. While interpretable, it is less so than BLR and NB.
- Gradient Boosted Decision Trees (GBDT) is another ensemble method that uses boosting-- combining a series of sequentially connected decision trees, each learning from the errors of the previous one—to produce a highly efficient and

accurate model. The strengths and weaknesses of GBDT are similar to RF; both include a feature importance estimate capability for built in feature selection.

1.5.4.3 Performance Evaluation

Using leave one case out cross validation, we will empirically assess and compare the predictive capabilities of the models. Leave one case out cross validation predicts each instance, training on all other (n-1) instances. An advantage of this approach is its ability to cover all data points and learn everything, reducing risk of model bias. However, this approach could be computationally expensive.

If computational burden becomes an issue, *k*-fold cross-validation, with *k*=20, will be considered. This evaluation method, while less computationally intensive, still covers a high number of data points, however risk of bias will be higher due to the potential of leaving out a positive case during an iteration. If we encounter model performance issues due to class imbalance, then we can consider using a stratified cross-validation approach.

1.5.5 Specific Aim 3b: Compare Model Capability to Predict Dynamically Evolving New-Onset HA-VTE.

Models will be assessed and ranked by performance metrics immune to unbalanced data: sensitivity (recall), specificity (TN rate), and precision (PPV). Area under the receiver operating characteristic (AUROC) curve will be used to quantify the tradeoff between sensitivity and specificity and will be compared with the area under the precision recall (AURPR) curve. Furthermore, consideration of clinical benefits and costs of the various models and at various

thresholds will be included in model performance evaluation. Models with the best combination of predictive ability and parsimony will be identified.

We have made the a priori decision to identify the best performing model in stage 1 (BLR, NB, RF, GBDT) based on the metrics outlined above.

1.6 Data Analysis Plan

Data analysis will be conducted using IBM SPSS software, version 26, R, and customized machine learning algorithms developed by Carnegie Mellon University (CMU) Auton Lab. Additional data review and visualization will be supported by NOBLE Coder [113] NLP software and Tableau.

1.6.1 Data Quality Diagnostics

Data quality will be examined prior to and as part of Aim 1 specific analyses to examine data distributions, identify invalid data, identify patterns of missing data and to evaluate relationships between different variables.

1.6.2 Descriptive Statistics

Detailed descriptive and exploratory analyses of each variable will be performed, to yield standard descriptive summaries (see Tables 3 and 4). For continuous variables, appropriate descriptive statistics, including graphical representation, will be computed to describe sample

characteristics and determine variable distributions. For categorical variables, frequency distributions and percentages will be computed and examined.

As all categorical variables (new onset VTE, gender, ethnicity and admission diagnosis) are nominal level of measurement, central tendency calculations will include only the mode and measures of dispersion (variability) are not indicated. For the continuous variables listed in Tables 3 and 4, central tendency will be described by the mean and dispersion as standard deviations for normally distributed data and medians and interquartile ranges (IQR) for non-normal data distributions.

1.6.3 Data screening procedures

Prior to performing planned analyses, statistical assumptions will be evaluated descriptively (skewness and kurtosis values) and graphically (frequency histograms, normal probability plots, and bivariate scatter plots) to check for violations in normality, linearity, and homoscedasticity of variables. (Table 5)

Table 5. Data Screening Process

1. Inspect univariate descriptive statistics for input accuracy Out of range values Plausible means and standard deviations Univariate outliers
2. Evaluate missing data for quantity and pattern Address problems found
3. Evaluate pairwise plots for linearity and homoscedasticity
4. Identify non-normal variables and univariate outliers and address accordingly Evaluate skewness and kurtosis

	Transform variables, if indicated/desired
	Check results of transformation
5. Identify and address multivariate (MV) outliers	
	Variables causing MV outliers
	Description of MV outliers
6. Evaluate variables for multicollinearity	

Adapted from Tabachnick, B. G., & Fidell, L. S. (2014). Using multivariate statistics. Harlow. [114]

1.6.3.1 Missing Data

As this research study employs data originally collected for clinical use, missing data is anticipated. Although we anticipate missing data will be missing completely at random (MCAR), patterns of missingness will be carefully and systematically evaluated. Deletion of cases or variable categories with missing values will be considered, as will imputation methods, as indicated by the missing data patterns found. However, many machine learning algorithms are robust to missing data and since EHR clinical data is often missing due to a variety of causes (sensor artifact, off-unit for tests), modeling approaches that can perform well in the setting of missing data are desirable for translation to clinical practice.

1.6.3.2 Outliers

Categorical variable frequency (maximums, minimums, and percentiles) will be examined for unbalanced distributions (i.e., categories < 10%) to determine outliers. Among continuous variables, univariate outliers will be screened using graphical methods (histograms, box plots, normality probability plots) and *z*-scores. Mahalanobis distance scores will be computed and screened to identify multivariate outliers. Discovery of outliers will be followed by a reexamination of data to assess for spurious values.

1.6.3.3 Data Transformations

If a normal distribution is not present, skewness direction will guide data transformation strategies. Winsorization of outliers by variable transformation (square root, logarithm, or inverse) and/or score alteration will be considered and implemented as needed. Outliers will be described, and all alterations and transformations will be reported. If transformations are indicated, all assumptions will be re-evaluated post-transformation.

1.6.3.4 Multicollinearity and Singularity

Variance inflation factors (VIF), tolerance, and Belsley, Kuh, and Welch diagnostics (condition indices) [115] will be used to assess for multicollinearity. VIF measures greater than 10 indicate serious multicollinearity and warrant close examination. Tolerance ($1-R^2$) values less than 0.3 will be investigated for large standard errors. Multicollinearity will be further assessed during regression analyses (5.2.3).

1.6.4 Data Analytic Procedures

1.6.4.1 Aim 1: Assemble a large scale, multi-source, multidimensional dataset from adult hospital step down unit (SDU) patients

Once static and time-stamped dynamic patient data have been linked to produce an intensive longitudinal dataset, descriptive analyses (1.6.2) and data screening (1.6.3) will be conducted as outlined above.

1.6.4.2 Aim 2: Ground truth case ascertainment of new onset, HA-VTE

The procedures for ground truth ascertainment of new-onset, HA-VTE and the annotation of VTE outcome events is detailed in section 1.3.2.1, section 1.5.2, and in Dissertation Manuscript #2 (Appendix B).

1.6.4.3 Aim 3: Develop and Evaluate Models to Predict New-Onset, HA-VTE Risk in Hospitalized Patients

An established predictive relationship between variables of interest is a prerequisite for ML models that provide robust and actionable insights [98] and that has been established in preliminary work (1.3.2.2 and 1.3.2.3). BLR will be the first classifier employed for the achievement of Aim 3 and it will serve as the base model for comparison. The relationship between static and dynamic variables and the binary dependent variables of interest will be examined. Assumptions specific to BLR that will be explored include:

1. Independence of observation. Identification of any duplications of subject information will be met in the achievement of Aim 1 (1.5.1) ensuring this assumption will be met.
2. Categories of the dependent variable and all nominal independent variables should be mutually exclusive and exhaustive.
3. Adequate sample size. This has been previously established in Section 1.4.3.
4. A linear relationship between the continuous independent variables and the logit transformation of the dependent variable. Linearity will be assessed vie the Box-Tidwell procedure. Residual plots of studentized residuals versus predicted values will be used to evaluate linearity. If a nonlinear relationship is discovered, transformations may be completed to enhance linearity, and appropriate regression analyses will then be performed.

5. Absence of significant multicollinearity. Multicollinearity will be assessed through an inspection of correlation coefficients and Tolerance/VIF values. If present, evidence-based relevance of patient factors will inform variable elimination.
6. No significant outliers, high leverage points or highly influential points. Outliers will be identified and addressed as outlined in 1.6.3.2 and 1.6.3.3 above.

Naïve Bayes, RF, and GBDT classifier algorithms do not require specific assumptions to be met and the selection rationale for the proposed algorithms has been previously summarized in section 1.5.4.2. Our outcome variable of interest is the minority class in our sample and this class imbalance has already been considered with our algorithm selection choices. However, if model performance is poor, additional considerations to address class imbalance in model building include:

1) **Under sampling of the majority cases.** This approach can prevent the signal of the majority class from dominating the learning algorithm. The disadvantage with under sampling is that it discards potentially useful data.
2) **Over sampling of the minority cases.** The disadvantage of oversampling is that by making exact copies of existing data, the risk of overfitting is increased.
3) **Synthetic Minority Oversampling Technique (SMOTE).[116]** This approach under samples the majority class and creates synthetic samples from the minor class, creating less risk of overfitting.

Application of ML algorithms and the process for comparing and evaluating performance of the subsequent models has been previously outlined in detail in Section 1.5.4 above.

1.7 Limitations

We expect Aim 1 to be the most challenging and time intensive aim and recognize Aims 2 and 3 rely on its achievement. Aim 1 builds on the applicant's pilot work and emerging skills with data annotation and curation and the ongoing guidance and assistance of bioinformatics expert consultant, Melissa Saul, ensures completion of this aim.

The parent sample is from a prior timeframe. Little has changed in diagnostic testing for VTE, or VS data accrued from bedside monitors and we do not anticipate this temporal gap to impact results. Every effort will be made to map clinical data from this sample to current data standards (such as mapping ICD-9 codes to ICD-10 codes) to maximize contemporary understanding and application.

We recognize the possibility that some of the "Never tested cases" may contain HA-VTE that is unrecognized by clinicians and that these unrecognized positive cases may not be diagnosed until after this hospital admission. We acknowledge this is a limitation and that exploring this limitation is beyond the scope of this current study. However, using the classifier trained for clinical rationale #2, we will explore its ability to predict VTE (+) cases from never tested cases and report class distributions, setting the foundation for future inquiry.

As this is the first study to develop prediction models for new onset HA-VTE using data accrued during hospitalization, we recognize that the time windows proposed may not be the most optimal time frames for predicting this dynamic disease pathology. They are, however, a starting point that will inform the aims of this dissertation study while also laying the groundwork for future lines of inquiry that will explore additional time windows in advance of diagnosis. Furthermore, our BLR model will allow us to compute confidence intervals that can inform future work exploring the cost-benefit-risk trade off to earlier detection times

Generalizability of findings will be limited due to the fact that the sample population is restricted to a single SDU over a specific time interval, and that sample ethnicity is primarily White. However, sample size allows for diversity of diagnoses, variables, and adequate VTE prevalence to support machine learning (1.4.3.1). This dissertation study is focused only on training and testing models using cross-validation methods. We recognize that HA-VTE models will need to be further tested and externally validated and while those steps exceed the scope of the dissertation research and the training plan time-line, they will be a focus of post-doctoral work.

1.8 Future Directions

The development of a new-onset, HA-VTE prediction models, will provide a valuable and novel contribution to the field of precision medicine science. Completion of this study will equip the applicant with experience and skills needed to conduct research as an independent nurse data scientist. Study results will serve as the foundation for the applicant's future research as a post-doctoral fellow which aims to 1) further assess model performance and generalizability on larger datasets from different geographic locations, health system sizes, and patient populations, 2) identify features discriminating between VTE subtypes (DVT versus PE), and 3) translate research findings to inform the design, implementation and testing of a dynamic VTE RAM to inform clinical practices for VTE prophylaxis and treatment.

1.9 Human Subjects Research Risks and Protections

1.9.1 Responsible Conduct of Research Training

Collaborative Institutional Training Institute (CITI) online training courses have been completed and include: Good Clinical Practice Course for Clinical Trials, CITI Conflicts of Interest, Biomedical Human Subjects Research, International research and Responsible Conduct of Research (RCR). Through completion of these modules, the candidate has become familiar with the general professional norms, accepted practices and ethical principles in the performance of all activities related to scientific research.

Additionally, completed School of Nursing (SON) coursework includes RCR topics including conflict of interest, plagiarism, data acquisition, scientific misconduct, authorship, presentation of data, informed consent, justification for randomized clinical trials, utilization of human and animal subjects.

A full list of RCR coursework can be found in Appendix H.

1.9.2 Risks to Human Subjects

This is not a clinical trial, but is a descriptive study using data from a retrospective time period. All patients in the retrospective study interval (11/06-9/08, total n=3864) are discharged from the hospital. De-identified vital sign, clinical, and demographic data (outlined in section 3.1) previously gathered through the parent study (R01NR014221) reside on a pre-existing secure study server. Only data that was de-identified using De-ID ™ software, are recorded in the study files, and study ID and visit ID are used for participant identification. The extraction of additional data

not previously collected but needed to support the proposed study that will augment the parent study data is limited to the parent project time interval. These de-identified data (radiology, ultrasound, and progress note data) will be obtained through Melissa Saul, MS, assigned a visit ID linking these additional data to the same patients' previously collected data, and will reside on the same secure server.

1.9.3 Human Subjects Involvement, Characteristics and Design

The patient population in the study is consistent with the characteristics of patients admitted to monitored beds on Unit 9G, University of Pittsburgh Medical Center (UPMC), Presbyterian Hospital over the study period 11/06-9/08. The racial, gender, and age characteristics of the subject population reflects the monitored patient population of the unit during the above study periods.

1. **Inclusion criteria:** Entry criteria were patient need for a monitored bed on Unit 9G (age > 21 years). Patients were admitted to the study unit according to the usual standard of care for monitored bed admission and utilization, and there were no special efforts to direct patient admission to Unit 9G, yielding a convenience sample of patients admitted to this unit and these monitored beds. Patients remained on Unit 9G, and a part of the study until their discharge from Unit 9G (discharge to another unit, discharge from the hospital).
2. **Exclusion criteria**: Patients with a preexisting diagnosis of VTE will be excluded, since the study purpose centers on new-onset VTE acquired during hospitalization as a FTR complication.
3. **Inclusion of special classes:** No special classes of patient on Unit 9G in the retrospective study interval, (women of childbearing age, pregnant women, prisoners and institutionalized individual) were excluded. Children were not included as participants in

the parent study and the entry criterion was set at > 21 years of age. UPMC Presbyterian is an adult care facility, with pediatric patients admitted to UPMC Children's Hospital. Any patients age < 21 years who were admitted to the study unit cohort spanning patient admissions between 11/06 and 9/08 were noted and their data eliminated from the analyses.

1.9.4 Protection Against Risk

Recruitment and informed consent. The VS, clinical and demographic data for all patents in the parent project interval were obtained and continue to be evaluated under active IRB approval. Use of these data for an expanded research agenda, using the previously collected data, augmented by unstructured clinical data, in a slightly different manner to answer a different research question, will now be supported under a separate IRB-approved protocol. Informed consent was waived for all data collection in the parent study, because:

- No identifiers were recorded by the study term.
- It would not have been possible to collect data on the full patient census and meet the aims of the study without the waiver.
- Patients were discharged from the hospital between 4-5 years prior to the data collection.

Only de-identified study data are stored on a password protected server that resides within the University of Pittsburgh School of Nursing's firewall. The only potential risk could be breach of confidentiality based on access to the study code linkage files. This risk is low as these are saved in the De-ID encrypted file.

1.9.5 Data Safety and Monitoring

The applicant and the Sponsor and Co-Sponsor (Hravnak & Pinsky) will be responsible for the ongoing evaluation of the progress of the research study. They will ensure that no patient Personal Health Information has entered the study database. During bi-monthly meetings, Drs. Hravnak and Pinsky will review progression of the study, data integrity, and preliminary results when available. Any breaches in data safety will be investigated and reported to the IRB. This study is a clinical study, but not a clinical trial, therefore, a separately uploaded data safety and monitoring plan is not required.

To summarize and reiterate: There is no risk of physical harm to the patient by being in the study. The only risk to the patient would be a remote breach of confidentiality. However, we have minimized the opportunity for that to occur by collecting only one identifier--the MRN--for linkage code purposes only, and even that is available only to the data extractor of the parent study. Once the clinical data elements were collected, they are maintained in a research file identified only by study ID.

1.9.6 Potential Benefits of the Proposed Research

This is a survey study, and there are no direct benefits to the patients.

1.9.7 Importance of the Knowledge to be Gained.

The leading cause of unplanned hospital death, the FTR complication of VTE, is associated with significant increases in health care costs, hospital lengths of stay and overall

mortality. The proposed study addresses a critical gap in current VTE risk knowledge and assessment approaches and has the potential to identify patterns of cumulative and complex patient factors predictive of VTE leading to a better understanding of the biological underpinnings of this common clinical complication. Knowledge generated from this project will inform more personalized feature patterns contributing to new-onset VTE in hospitalized patients translating to more accurate patient risk stratification, earlier identification of complications and guiding interventions to inform nurse-driven clinical decision support at the bedside, and improve nurse-sensitive patient outcomes.

1.9.8 Institutional Review Board (IRB) Approval

This study was approved by the University of Pittsburgh IRB on June 14, 2018 and approval of continuing review was granted on June 17, 2019. IRB approval documentation can be found in Appendix H.

2.0 Summary of Study

2.1 Study Overview

The purpose of this dissertation study was to develop and evaluate models predictive of the failure to rescue (FTR) complication of dynamically evolving hospital acquired venous thromboembolism (HA-VTE) event risk in hospitalized patients, using progressively granular, intensively collected time series data. This dissertation topic and study grew from complexities, questions, and challenges surrounding FTR patient complications that the candidate encountered in her clinical practice as a nurse and nurse practitioner.

Dissertation Manuscript #1, "ICU Scoring Systems," has been accepted for publication in *Critical Care Nurse.* This manuscript was developed through review of the literature exploring tools available to clinicians at the bedside to enable clinical prognostication and risk assessment of complex patients and the abstract is included in Appendix A. This literature review provided the candidate with deep foundational knowledge of the strengths and limitations of clinical risk prediction scoring systems and how these systems are developed, validated, and utilized in clinical settings. Additionally, the need for routine reassessment and revision to stay current with clinical practices and specific populations was highlighted.

Three preliminary/pilot studies were conducted (detailed in Section 1.3.2) that established ground truth prevalence of HA-VTE cases in the study sample and a significant relationship between the predictor variables and HA-VTE outcome, supporting the premise of this proposal. These preliminary analyses have been disseminated in two oral conference presentations and three published abstracts [95, 98, 99] and contributed to the completion of Aims 1 and 2.

An overview of study methods is provided in Figure 12. The results for Aim 2 comprise Dissertation Manuscript #2, "Accuracy of Identifying Venous Thromboembolism by Administrative Coding: Implications for Big Data and Machine Learning Research," which has been accepted for publication in the *Journal of Clinical Monitoring and Computing*. The abstract of this manuscript is included in Appendix B.

The results of dissertation Aim 3 are presented in the databased manuscript, "Predicting Dynamically Evolving New-Onset Venous Thromboembolic (VTE) Event Risk in Hospitalized Patients," and included in section 3.0 of this document. This will constitute Dissertation Manuscript #3 to be submitted to the *American Journal of Critical Care*.

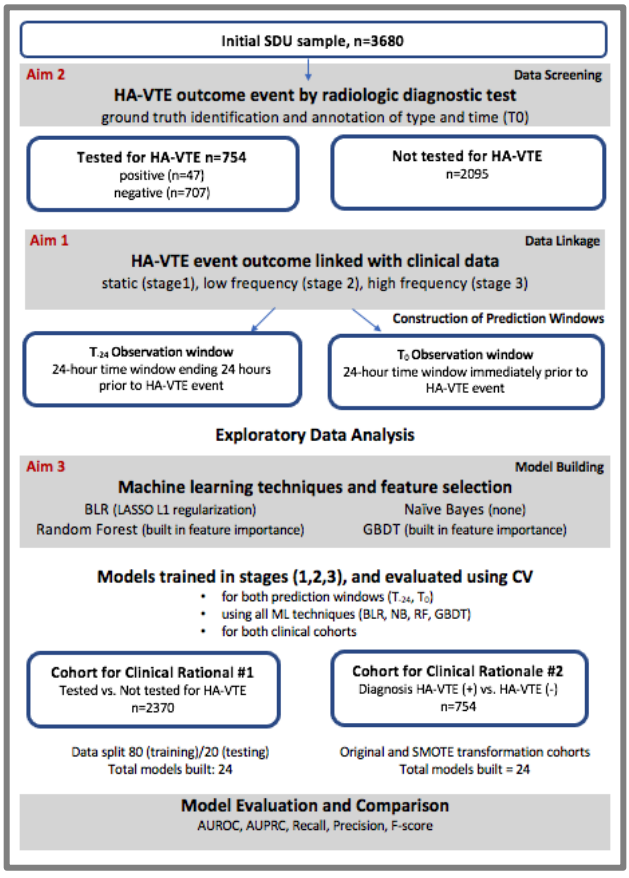


Figure 12. Overview of Study Methods

2.2 Strengths and Limitations

This hypothesis driven study is the first to formally incorporate the progressive accrual of multi-domain, granular, intensively collected time series data during hospitalization in the development of models to predict risk of the dynamic complication of new onset, HA-VTE. Additionally, it is the first study to endeavor to predict risk within specific time windows in advance of HA-VTE diagnostic test/diagnostic test results. Our foundational models not only demonstrate the addition of dynamic clinical data improves HA-VTE risk prediction, but they also provide emerging insights on how clinical features associated with evolving disease pathology change over time.

Quality machine learning (ML) methods are predicated on data veracity. The quality of data used, and the rigor of data screening and ground truth annotation are significant strengths of this study. Collecting continuous VS data in patients outside of the ICU or operating room is a challenge and having data of this level of granularity for medical-surgical hospital patients is rare. Many prior VTE risk studies are limited in that they report lack of confidence that their control cohorts are free of unidentified VTE cases and the rigorous and reproducible process of HA-VTE positive and confirmed negative case identification (Aim 2) by radiologic diagnostic test review is another strength of this work.

The ability to develop models that can be translated to clinical practice necessitates constructing models based on data sets that reflect real world population incidence of VTE disease, which varies from 1-16% in medical patients to as high as 40% in certain surgical and oncologic populations. The low prevalence of HA-VTE used in our modeling approaches is a strength as it reflects prevalence of HA-VTE in the population.

Despite the strengths of this study, there are important limitations that should be acknowledged. While our sample size allows for diversity of diagnoses, variables, and adequate VTE prevalence to support machine learning, we recognize that generalizability of findings will be limited due to the fact that the sample population is restricted to a single SDU over a specific time interval, and that sample ethnicity is primarily White.

The high frequency VS data used in model development was only available during the time that patients in this sample were in the SDU and thus our ability to explore HA-VTE risk using these dynamic data for modeling, was limited to this time frame. An extensive list of clinical prediction variables was proposed, based on literature review and mapping to published and commonly used VTE risk assessment tools. While most of those variables were technically represented in the data, the prevalence of those clinical feature variables in our sample population and during SDU stay was low. As a result, many variable categories were excluded due to absence of information that would only add noise and bias to model development. This reality has provided valuable insights that will guide data collection strategies and rationales in future work.

Venous thromboembolism is a disease process that evolves slowly, varies by individual and pin-pointing the exact time VTE pathology begins is as difficult in retrospective data as it is in the clinical setting. When clinicians suspect VTE pathology, diagnostic tests are ordered and conducted quickly so that appropriate interventions are initiated without delay. We determined the time that the VTE gold standard diagnostic radiology test was conducted to be the best proxy for HA-VTE event outcome determination in this study. A limitation of this approach is that some of our positive cases may have had VTE pathology present for hours to days in advance of diagnosis. Ideas for future work include developing models with annotated outcome event defined as the time the gold standard radiologic diagnostic test for HA-VTE is ordered by providers, to better

approximate clinician suspicion of disease presence. Additionally, we recognize the possibility that some of the "not tested cases" may contain HA-VTE unrecognized by clinicians and that these positive cases may go unrecognized without testing, or not be diagnosed until after this hospital admission. Exploring this limitation is beyond the scope of this current study, however, exploring this constraint is a consideration for future work.

As this is the first study to develop prediction models for new onset HA-VTE using data accrued during hospitalization, the time windows proposed may not be the most optimal time frames for predicting this dynamic disease pathology. Predicting HA-VTE at time of diagnosis, while not meaningful for applying preemptive clinical intervention to avoid the complication, is a requisite first step in temporal risk modeling research that aims to identify patient features associated with the evolution of complication development and risk. The chosen prediction times represent an initial starting point that informed the aims of this dissertation study while also establishing the groundwork for future lines of inquiry that will explore additional time windows in advance of diagnosis.

Finally, this dissertation study took a survey approach to model development. A variety of classification algorithms were employed, to explore model development with a variety of sample cohorts, prediction time windows, and temporal clinical data. However, we focused only on training and testing models using cross-validation methods. Future work should include 1) determining the optimal set points of individual models, 2) hyperparameter tuning to enhance model performance at chosen set points, and 3) evaluating model performance with an external data set. While those steps exceed the scope of the retrospective dataset attributes available for analysis, dissertation research and the training plan timeline, they will be an ongoing priority in

the candidate's post-doctoral work and foundational to her research trajectory which is focused on predicting FTR complication risk to inform targeted prevention and treatment decisions.

2.3 Future Studies and Implications for Nursing

This study provided important doctoral research and methodologic training and produced findings that establish a foundation for future lines of inquiry. Using the classifier trained for clinical rationale #2, we will explore its ability to predict HA-VTE (+) cases from never tested cases and report class distributions. Future work will also include fine tuning model parameters to improve model performance, evaluation using "never-before-seen" data, and replication of this study with an expanded data set that includes greater diversity and higher prevalence of the predictor variables of interest as well as the HA-VTE outcome. Important future directions of this study include exploration of different time windows in advance of event outcome then comparison of model performance with existing VTE risk assessment models.

Surveillance is a nursing intervention critical to patient and safety outcomes that involves the purposeful and ongoing acquisition, interpretation and synthesis of patient data to inform clinical decision-making. A prerequisite to timely intervention, nursing surveillance can be affected by a variety of factors: nurse education level, clinician experience, clinician fatigue, alarm fatigue, workplace processes, staffing ratios and unit skill mix. Notably, inadequate surveillance is strongly associated with increased FTR rates. This dissertation study is the first to exploit ML approaches using high frequency VS data, to inform the critical practice of nursing surveillance for a nurse-sensitive patient complication and it has strong potential for future clinical translation.

Machine learning models can provide robust and actionable insights that can better inform nursing surveillance practices (needed frequency of monitoring, staff allocation), clinical decision making (timely and accurate recognition, treatment selection), and care delivery systems (patient triage, diagnostic testing, adverse event prevention). Our initial findings have allowed us to demonstrate the value of incorporating progressively accruing, routinely available, continuously acquired clinical data in the prediction of risk for complications that evolve slowly over the course of hospitalization. This is especially true for the preventable complication of VTE, for which current risk assessment guidelines focus on static clinical variables and can include expensive and specialized serum studies and/or extensive compilation of medical history variables, which is not always possible given the constraints of patient acuity or competing demands for clinician time.

This study aligns with NINR's strategic vision for nurse scientists to employ new strategies for collecting and analyzing multi-dimensional data sets to permit better understanding of the biological underpinnings of health and improve ways nurses prevent and manage illness. During the conduct of this dissertation work, significant insight was gained into the importance of training nurses in clinical data extraction, storage, and annotation, in basic *and more advanced* statistical programming languages, bioinformatics, and in building interdisciplinary working relationships with nurses and computer scientist experts during doctoral training. Developing high quality predictive models that can meaningfully impact clinical practice and patient outcomes requires close and constant cooperation between clinicians and computer scientists. It is recommended that nursing PhD students interested in data science methodologies should receive, at a minimum, hands-on introductory training in statistical programming and machine learning methodologies, as well as interdisciplinary coursework and practicum training that is specifically designed to include a mix of nursing, medical, computer science, biomedical engineering, and biomedical informatics

students. A strong understanding of the technical language used by different domains and of how the collaborative intersection of expertise of different disciplines is integral to data science focused on health care outcomes, should be a foundational component of nursing PhD training.

Nurses are uniquely positioned on the frontlines of care. No other healthcare worker spends as much time at the bedside *during the course of routine clinical care*, collecting and recording physiologic data, tracking symptoms, interfacing with patients, families, and other members of the health care team, and providing advocacy and health promotion education. Thus, the lens of nursing expertise is very much needed in data science and predictive analytics research to provide important insights into clinical workflow practices that are critical to the development of research design, data annotation, and data interpretation. The value to nursing research and practice, of concretely promoting interdisciplinary training and data science skills in nursing science doctoral education is important and cannot be overstated.

3.0 Data Based Manuscript: Differentiating Dynamically Evolving New-Onset Venous Thromboembolic (VTE) Event Risk in Hospitalized Patients with Machine Learning

3.1 Abstract

Background. Hospital acquired (HA) venous thromboembolism (VTE) is the leading cause of preventable hospital death. VTE pathology and symptoms evolve slowly over hours to days and prophylaxis with anticoagulant medications carries risk of bleeding. No current HA-VTE risk assessment models incorporate the progressive accrual of dynamic patient data over time of hospitalization. Classification algorithms which incorporate prediction time windows hold promise for closing this gap.

Methods. An observational, retrospective, cohort study (n=2370) was conducted to develop predictive models to classify patients at risk for HA-VTE during SDU admission. Expert review of gold standard diagnostic radiology tests identified ground truth HA-VTE outcomes (+/-, not tested) and pre-existing cases were excluded. A suite of classification algorithms consisting of binary logistic regression (BLR), naïve Bayes (NB), Random Forest (RF), and Gradient Boosted Decision Tree (GBDT) were used to train models for two prediction time windows: 1) hours 48 to 24 in advance of, and 2) hours 24 to 0 in advance of HA-VTE event time (T_0), defined as time gold standard diagnostic radiology test was conducted. Performance was evaluated with 10-fold stratified cross-validation. Two model sets were selected to illustrate two relevant clinical scenarios. To help inform risk stratification and identify patients in need of closer surveillance, the first set of models (S+/-) differentiated between patients in whom clinicians suspected HA-VTE

and ordered diagnostic radiology evaluation (n=760) from those not suspected and not tested (n=1614). Among the patients identified as higher risk, and tested by clinicians, a second set of models (C+/-) was developed to differentiate between confirmed positive (n=47) and negative (n=713) diagnostic test results. Both S+/- and C+/- model sets were built using a stage-wise process that added data with increased granularity in 3 successive stages: Stage 1) only data present-on-admission; Stage 2 low frequency medication and laboratory data added; and Stage 3) added high frequency vital sign data, collected at a rate of once every 20 seconds from bedside monitors. Performance was evaluated for each model set at each stage using metrics robust to class imbalance and prioritizing recall (TPR).

Results. All models demonstrated improved precision-recall performance with progressive addition of dynamic clinical data. For S+/-, NB, with a TPR of 18% (Stage 1, static), 73% (Stage 2, LF added), and 77% with HF data added in Stage 3, was the most *sensitive* model for classifying at T_0 prediction time point, (AUPRC of 0.56, a PPV of 59%, AUROC of 0.66). The RF model with a TNR of 63% (Stage 1, static), 78% (Stage 2, LF added), 81% with HF data added in Stage 3, was the most *specific* model for classifying at T_0 prediction time point (AUPRC of 0.60, PPV of 62%, AUROC of 0.70). When using dynamic LF and HF data accrued 48 to 24 hours in advance of the event to classify cases at a prediction time 24 hours in advance of T_0 (T_{-24}) the NB model TPR was 76% (AUPRC .52, PPV 46%, AUROC .60) and the RF and GBDT models identified true negatives with a specificity of 80%.

For the C+/- models, including only patients identified at higher risk for HA-VTE, BLR with a TPR of 29% (Stage 1, static), 66% (Stage 2, LF added), and 94% with HF data added in Stage 3, was the most *sensitive* model for classifying at T_0 prediction time point (AUPRC of 0.81, PPV of

84%, AUROC of 0.90). The RF model with a TNR of 94% (Stage 1, static), 96% (Stage 2, LF added), 97% with HF data added in Stage 3, was the most *specific* model for classifying at T_0 prediction time point (AUPRC of 0.97, PPV of 96%, AUROC of 0.93). When using dynamic LF and HF data accrued 48 to 24 hours in advance of the event to classify cases at T_{-24}, the NB model had a 91% TPR (AUPRC .77, PPV 53%, AUROC .68).

Clinical features of heart rate, respiratory rate, dynamic hematologic labs, medications, and trauma admission were identified as most important in the Stage 3 models.

Conclusion. Classification algorithms applied to routinely collected dynamic clinical data improve HA-VTE risk prediction ability over static data models and have the potential to improve detection of at-risk patients.

s.

3.2 Introduction

Failure to rescue (FTR) is a national metric of care quality that represents the death of a hospitalized patient due to a treatable complication arising after hospital admission.[1] Initially associated with surgical patients, FTR is now recognized as a multi-faceted phenomenon that can occur in any hospital setting, from a variety of both surgical and non-surgical causes. [2]

Hospital acquired venous thromboembolism (HA-VTE), a failure to rescue (FTR) complication manifesting as deep vein thrombosis (DVT) or pulmonary embolism (PE), is the leading cause of preventable hospital death, carrying a high mortality risk and a national cost burden of $7 billion annually.[7-9] Clinical signs and symptoms of evolving venous thromboembolism (VTE) are subtle, presenting gradually over hours to days. Reasons cited for the persistence of this major public health problem include: 1) complexity of known VTE risk factors and unknown gaps in VTE risk knowledge; 2) risks and limitations of VTE prophylaxis and; 3) limitations of current VTE risk assessment models.[37, 40]

3.3 Background

The pathology of HA-VTE is complex, involving interactions between clinical risk factors and acquired and/or inherited susceptibilities to thrombosis. Rates of VTE risk range from 1% in medical patients to as high as 40% in certain surgical and oncologic populations.[9, 36] Some risk factors are associated with an increased overall lifetime risk (older age, venous insufficiency, obesity) and others with provoking acute risk (malignancy, recent surgery, indwelling vascular catheters). [9, 47-52] Multiple risk factors are hypothesized to synergistically increase patient risk,

but the relative contribution of each factor to overall VTE risk is unclear. Symptoms of new-onset VTE often occur gradually over a period of hours to days and clinicians can easily fail to notice subtle feature patterns until after a critical event occurs.[64, 65] Individual symptoms and risk factors for VTE have low predictive value (about 15%),[67] underscoring the need to identify dynamic composite feature pattern evolution associated with VTE.

VTE prevention guidelines vary across patient populations, making consistent implementation a challenge for providers.[68, 69, 71] Routine prophylaxis of VTE with low molecular weight heparin (LMWH) is a standard recommendation for hospitalized patients; however, it is not a panacea for prevention. Indiscriminate anticoagulant prophylaxis in patients without significant risk factors is associated with impaired wound healing, increased bleeding, need for blood transfusions and increased patient discomfort costs.[71, 117] Conversely, underutilization of VTE prophylaxis has been linked to provider fear of these harms.[68-73] Nonpharmacologic prophylaxis methods, such as elastic and pneumatic compression stockings, are lower risk prevention strategies. However they can be uncomfortable for patients and are inconsistently implemented in hospital settings.[118] Identifying patient feature patterns predictive of VTE risk can inform more precise risk stratification and prophylaxis application, helping those patients who need it most, but limiting iatrogenic complications in those who do not.

Current VTE risk assessment models, the cornerstone of prevention, have limited utility due to their complexity, lack of reliability, generalizability and external validation, and dependency on static data. [10, 11] Some models include specialized coagulopathy or genetic serum studies not routinely ordered on all patients.

VTE risk is associated with interactions between established and acutely acquired risk factors and pathology and symptoms evolve over the course of hospitalization. Recommendations

for optimal VTE risk assessment models, published by the Agency for Health Care Research and Quality, emphasize the need for tools that are 1) simple to use in routine clinical practice, with minimal need for specialized laboratory studies, complex calculations, or extra charting, 2) lend themselves to automation using information readily available in the EHR, and 3) that incorporate dynamically accruing data and ongoing reevaluations.[119] However, current RAMs share a common and significant limitation: they rely primarily on static, baseline patient features to assess risk for VTE. A critical gap in VTE risk modeling research is that while VTE risk and pathology evolves over the course of hospitalization, no current models incorporate the progressive accrual of dynamic patient data and pattern evolution over time in their modeling approaches.

HA-VTE risk changes over time-course of hospitalization and exploring dynamically accruing clinical data holds potential for informing this gap. Machine learning (ML) methodologies have the ability to scale up correlational analyses and discern emerging complex patterns and relationships. This approach has been used to successfully develop highly sensitive models to predict risk of cardiorespiratory instability, cardiac ischemia and arrhythmia, [80, 81, 84, 85, 120] and inform decision making to improve care delivery and patient outcomes. The objective of this study was to apply classification algorithms to dynamic clinical data (patient data marked by activity or change during the hospital admission) readily available to clinicians and routinely collected over the course of a step-down unit admission LOS, to develop and evaluate models for new-onset, HA-VTE risk.

Using a three-stage approach that incorporated time series clinical data increasing in collection frequency with each stage, we developed two sets of classifier models informed by the following clinical rationales:

1. **Clinical Rationale 1.** To determine ability to identify patients who clinicians identify as high risk for HA-VTE, we explored the following question: Among all SDU patients, can we discriminate between cases clinicians suspect of HA-VTE pathology and order a radiologic VTE diagnostic test from those they do not suspect, and do not test? (S+/- model sets, prediction outcome variable is HA-VTE diagnostic test: tested for HA-VTE versus not tested)
2. **Clinical Rationale 2**. To determine ability to identify which high risk patients will develop a confirmed HA-VTE diagnosis, we explored the following question: Among patients for whom clinicians ordered a VTE radiologic diagnostic test, can we predict confirmed positive diagnoses from negative diagnoses? (C+/- model sets, prediction outcome variable is HA-VTE diagnostic test result: positive diagnosis versus negative diagnosis).

For each set of models, we endeavored 1) to predict HA-VTE risk at two prediction time windows (PTW) (between 48 to 24 hours before event [$PTW_{48\text{-}24}$ or $T_{\text{-}24}$ models], and between 24-0 hours before event [$PTW_{24\text{-}0}$ or T_0 models]) and 2) to determine if the sequential addition of dynamic clinical data of increasing granularity improved model performance.

3.4 Methods

3.4.1 Sample and Setting

The study sample was obtained from data previously collected for the Predicting Patient Instability Non-invasively for Nursing Care (PPINNC) study (R01 NR01391). Entry criteria for the parent study were patient need for a monitored bed in a university medical center SDU and age >21 years between 11/06 and 09/08. Under Institutional Review Board (IRB) approval for waiver for informed consent, every patient admitted during the study timeframe to the study unit contributed to the data without exclusion except for age <21, and no special classes of patients were excluded, yielding a convenience sample of 3680 patients. Data collection for each patient spanned the entire duration of the patient's SDU length of stay (LOS), for the total unit census across the study time frame. Patients discharged or transferred to a non-study unit were discontinued from the study, but their data prior to that time retained

As this study aimed to develop models predictive of new-onset, HA-VTE, patients with a diagnosis of chronic VTE were excluded. Also excluded were patients with new-onset, HA-VTE identified during the hospitalization but before SDU admission or >24 hours after SDU discharge, since the continuous VS data streams collected only on the study unit and necessary for the modeling were not available for those patients. The rigorous, ground truth VTE outcome event diagnosis identification methods have been previously detailed and published in abstract form.[95] Based on that review of gold standard radiologic VTE diagnostic tests (lower extremity Doppler ultrasound [LEDUS], computed tomographic angiography [CTA], ventilation-perfusion [V/Q] scan and/or magnetic resonance angiography [MRA]), patient cases were identified as never tested

for VTE (n=1614), tested for VTE with negative findings (n=713), and tested for VTE with positive findings (n=47), during their SDU LOS.

During the study time frame, some participants were admitted to the SDU more than once during separate hospital admissions and/or during the same hospital LOS. For this analysis, each SDU admission LOS (and all clinical and VS data streams during that admission) was considered a case thus, some study participants have contributed to more than one case used in analysis. As this analysis seeks to determine clinical patient features over time of SDU admission that are predictive of new-onset, HA-VTE, this approach is consistent with the study aims and does not result in any overlapping data use.

3.4.2 Variables

A list of all variables considered for use in this analysis and notation of how they were identified and extracted from the clinical data can be found in Tables 3, 4, 5, 6, and 7 of the ETD documents. Diagnoses not present for a patient, and laboratory results, medication doses, and/or procedures that were not present for a patient, in either clinical or charge data, were assumed to be *absen*t (meaning normal or not indicated) rather than unknown or missing.

3.4.2.1 Outcome Variables

Following patient exclusions as described, all remaining cases were annotated for one of the following HA-VTE outcome event categories detailed in Figure 13. and defined below:

HA-VTE outcome: Not tested. SDU patients for whom clinicians did not suspect VTE, and therefore did not undergo a VTE diagnostic test during their SDU stay or during the 24-hours

immediately after SDU discharge, were annotated as a not-tested case. Their hypothetical outcome event time (T_0) is defined for these cases by protocol detailed in the data preparation section below.

HA-VTE outcome: Tested. SDU patients who clinicians suspected VTE pathology risk and ordered a gold standard radiologic test during their SDU stay or during the 24-hours immediately after SDU discharge, were annotated as a tested case. This outcome category of tested SDU patients was further subdivided by test result outcome:

- **HA-VTE outcome**: **tested for VTE with positive radiology test results** includes both new-onset DVT and PE diagnoses. New-onset DVT cases were defined as an acute venous thrombosis in at least one of the following deep veins on LEDUS: internal jugular, superior vena cava, inferior vena cava, iliac, femoral, popliteal and profunda femoris veins. Chronic thromboembolic disease and thrombi in the following vessels were excluded from the definition of new-onset DVT: Portal circulation veins, superficial veins, man-made venous conduits and arteries. New-onset PE was defined as the presence of an acute occlusive or sub-occlusive clot in a main, lobar, segmental and/or sub-segmental pulmonary artery on CTA or MRA. Time stamp for the occurrence of a positive HA-VTE outcome event is annotated as time zero (T_0), representing the date and time the confirmatory diagnostic test was conducted. Indeterminate radiology test results subsequently confirmed as a new-onset VTE event, based on review of additional medical notes, were time stamped at the time of the indeterminate test.
- **HA-VTE outcome**: **tested for VTE with negative radiology test results.** SDU patients with a gold standard radiology test negative for the presence of embolus/emboli in the previously specified vasculature were annotated as confirmed VTE negative cases (1.3.1.2). Time stamp for the occurrence of a confirmed VTE negative outcome event (T_0)

is the date and time the confirmatory negative diagnostic test was conducted. Indeterminate radiology test results subsequently confirmed as negative for VTE, based on review of additional medical notes, were annotated for T_0 at the time of the indeterminate test.

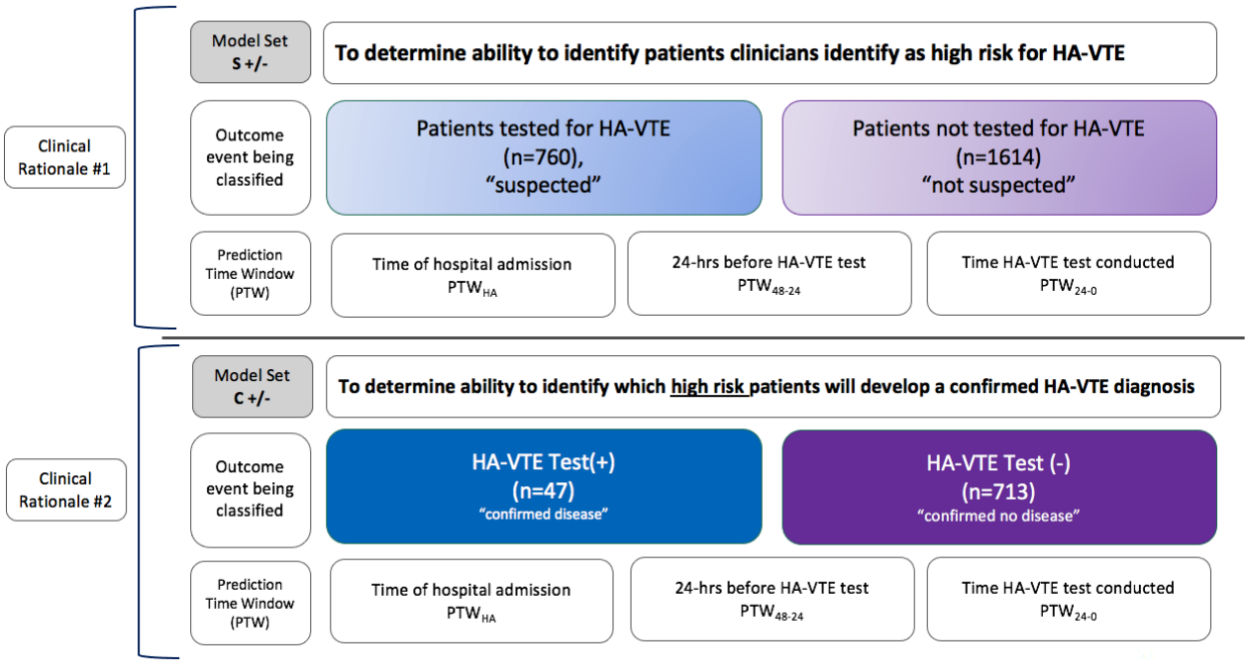


Figure 13. Clinical Rationale, Outcome Events and Prediction Time Windows for Each Model Set

3.4.2.2 Predictor Variables.

Figure 14 illustrates the three categories of independent variables were used in the stage-wise model development: Stage 1: Static data present upon admission, Stage 2: Accruing low frequency data, and Stage 3: Accruing high frequency continuous vital sign monitoring data. Predictor variables were identified based on review of the literature, cross-mapping of available EHR data to existing VTE RAMs, VTE disease pathology, and their availability in routine clinical workflow in advance of the outcome diagnosis.

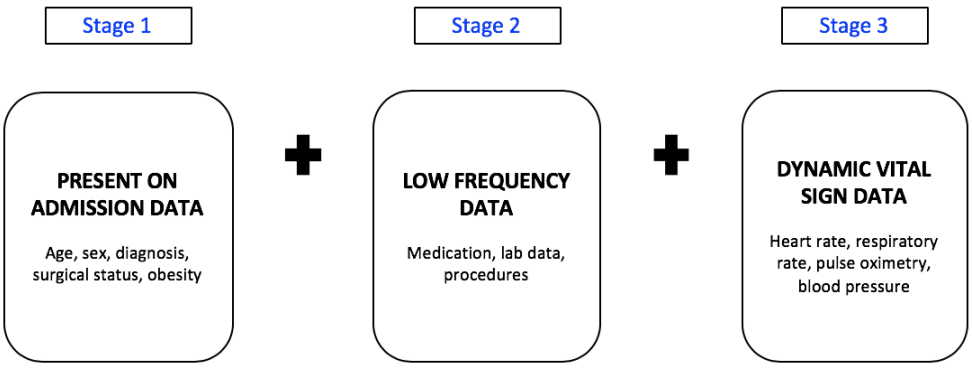


Figure 14. Review of Model Building Stages

Stage 1: Static data present on admission. Data readily available to clinicians at the time of hospital admission included the demographic information of age, race (White, Black, other, and unknown), obesity, VTE risk factor diagnosis(es) (based on published VTE RAMs) known at admission and Charlson Co-Morbidity Index Deyo (CCI-D) method scores. CCI-D score, based on clinical data from the current admission, is the only variable not technically comprised of data collected in advance of the outcome event we aim to predict. However, CCI-D scores for this population were compared with prior scores for a subset of the sample population and were not found to be significantly different. Thus, the decision was made to include this variable as a surrogate indicator of overall patient co-morbidity burden at admission. Mean CCI-D scores for this sample are notably low. A total of 29 variables comprising Stage 1: Static data present on admission were extracted for use in model building.

Stage 2: Dynamically Accruing Low frequency (LF) data. These data were accrued after SDU admission within a time granularity of hours to a day. Doses of medication by category, detailed in Appendix C, Table 9, (time granularity of calendar day) and laboratory test results (time granularity of date and time tests are resulted) are the LF variables included in model development. The time granularity of these data was constrained by the manner in which they were collected for

and available from the parent study database. Central venous catheter presence and surgical procedure data were identified as important variables and extracted from the MARS (Table 7, Appendix D) but ultimately excluded because of low prevalence within this population. A total of 235 LF clinical data variables were extracted and featurized for use in model building.

Stage 3: Dynamically Accruing High frequency (HF) data. Time stamped VS data containing heart rate (HR), respiratory rate (RR), peripheral capillary oxygen saturation (SpO_2) data collected at 20 second intervals (1/20Hz) from Philips bedside monitors, and non-invasive blood pressure (BP) was available when cycled, less frequently, but at a minimum of once every two hours. These data were originally collected for the parent study (total of 172,000 monitoring hours) and stored on a secure server. A total of 122 VS features (detailed in Tables 5-7 below) were extracted and available for use in model building.

Table 5. Static Present on Admission Data

Feature	EHR data point	Data type
Age	Patient DOB	Continuous Categorical
Gender (male/female)	Sex	Nominal
Prior VTE (yes/no)	ICD-9 code	Nominal
Known clotting disorder (yes/no)	ICD-9 code	Nominal
Obesity (yes/no)	ICD-9 code	Nominal
Malignancy (yes/no)	ICD-9 code	Nominal
Acute MI (yes/no)	ICD-9 code	Nominal
CHF (yes/no)	ICD-9 code	Nominal
Stroke (yes/no)	ICD-9 code	Nominal
SCI (yes/no)	ICD-9 code	Nominal
LE/Fracture (yes/no)	ICD-9 code	Nominal
Trauma (yes/no)	ICD-9 code	Nominal
Varicose veins (yes/no)	ICD-9 code	Nominal
Chronic venous insufficiency (yes/no)	ICD-9 code	Nominal
Acute infection/sepsis (yes/no)	ICD-9 code	Nominal
Charlson Co-Morbidity Index Deyo Method	ICD-9 code	Continuous

Table 6. Dynamic Low Frequency (LF) Data

Feature	EHR data point	Data type	Proposed Features
Central venous catheter	Charge data	Nominal	Yes/No
Major surgery (> 45-minute case)	DRG code	Nominal	Yes/No
Minor surgery (< 45-minute case)	DRG/procedure codes	Nominal	Yes/No
Joint replacement surgery	DRG code	Nominal	Yes/No
Medication data as per Table 6, Appendix C	Medication data	Nominal Ratio	Med ordered (Yes/No) Number of doses ordered
Lab data as per Table 4, Section 1.4.4.2	Lab data	Ratio	• Count • Number abnormal high results • Number abnormal low results • Max value • Min value • Mean value

Table 7. Dynamic High Frequency (HF) Data

Feature	EHR data point	Data type	Proposed features for each time window
Diastolic blood pressure (DBP)	Vital sign (VS)	Ratio	• Maximum value • Minimum value • Mean • Range • Standard deviation • Number of times measured during the time window
Systolic blood pressure (SBP)	VS	Ratio	
Mean arterial blood pressure (MAP)	VS	Ratio	
Heart rate (HR)	VS	Ratio	
Respiratory rate (RR)	VS	Ratio	
Pulse oximetry (SpO_2)	VS	Ratio	

3.4.3 Data Set Construction

Radiology reports (n=4544 reports) for lower extremity Doppler ultrasound, computed tomographic angiography, ventilation-perfusion scan and/or magnetic resonance angiography from our SDU convenience sample (n= 3680 cases) were extracted from the UPMC medical archival retrieval system (MARS)[92] and reviewed by a clinical expert to identify new-onset VTE positive cases and VTE confirmed negative cases. For all tested patients, the time (in minutes) from SDU admission to the time gold standard VTE diagnostic tests were conducted, T_0, was determined and defined as time-to-test. A hypothetical outcome event time (T_0) was defined for non-tested cases by the following process. To account for LOS (a well-established risk factor for HA-VTE), not tested cases were matched with tested cases based on SDU LOS. To enable evaluation of both tested and not tested patients at similar time points in the SDU stay, not tested patients were matched to the mean time-to test distribution for all tested patients and then a random sample from that distribution was obtained.

Tested (VTE confirmed positive and negative) cases and not tested cases were then linked with Stage 1 static data, and then with Stage 2 LF and Stage 3 HF data (Figure 12) for two prediction time windows (PTW) preceding the HA-VTE event, $PTW_{48\text{-}24}$ (between 48 to 24 hours before T_0 or T_{-24} models), and $PTW_{24\text{-}0}$ (between 24-0 hours before T_0 or T_0 models), producing a final sample of 2370 cases.

3.4.3.1 Data Screening

Exploratory data analysis was conducted, using IBM® SPSS® software version 27 (IMB Corporation, Armonk, NY) and R version 4.0.2 (R Core Team, 2020)[121] to examine data distributions, identify invalid data, identify patterns of missing data and to evaluate relationships

between different variables. As this analysis employs data originally generated for clinical use, missing data was anticipated, and patterns of missingness were systematically evaluated and found to be missing completely at random (MCAR). Missingness in static and LF data was < 1.0% and 5 cases (3 tested with HA-VTE negative results and 2 not tested) missing all demographic data, were excluded from analysis. HF vital sign data missingness ranged from 0.5% to 28.6 % depending on the variable. As there were significant differences in mean VS values between the HA-VTE outcome event classes, imputation using predictive mean matching (R, Mice 3.12.0),[122] was performed by class, to ensure those differences were not diluted. Assumptions and relationships were then re-evaluated.

3.4.3.2 Procedure for Identification of Predictor Variables Used in Modeling

Variables identified as germane to HA-VTE risk (based on theory and prior evidence) and available for this patient sample, were determined using criteria informed by Leisman et al.'s Guidelines for the Development and Reporting of Prediction Models.[123]

Availability in clinical workflow. Variables not easily obtained in routine clinical workflow or in advance of the outcome diagnoses, were excluded, with the exception of the CCI-D score, as previously mentioned.

Selection of Final Feature Set for Modeling. The list of variables developed from review of the literature (Appendix E) was then further pruned based upon availability in the data set, as well as relationship to HA-VTE event outcome variables. Once the data set was operationalized, we found low availability or presence of some variables in the Static and LF data categories**.** When clinical values were absent, and cross checking against charge data confirmed they were not present because they had not been ordered, we identified these data as absent. Absent values were present in large quantity in medication and lab data. Variable categories with data value absence

> 49% and those discovered to not be significantly associated with the outcome variables (defined in 4.3.2.1) in univariate testing were excluded from inclusion in model development. The subset of variables surviving these exclusions as described, were the features ultimately used in modeling and are listed in Table 5. These features next underwent statistical feature selection approaches as specified for each ML algorithm used, and the process is detailed in the following section.

Table 8. Features Used in Model Development

Features k= 65	Model Stage	Data type and number of features
Age, year **Sex,** (male/female) **Race,** (White, Black, other, unknown) **Charlson Comorbidity Index Deyo Method (CCI-D)**, score **VTE risk factor: Hypertension,** present/absent **VTE risk factor: Diabetes,** present/absent **VTE risk factor: Chronic Pulmonary Disease,** present/absent **VTE risk factor: Trauma,** present/absent **VTE risk factor: Diabetes Mellitus,** present/absent **VTE risk factor: Ischemic Heart Disease,** present/absent **VTE risk factor: Septicemia,** present/absent **VTE risk factor: Congestive heart failure,** present/absent	Stage 1	Static k=12
Narcotic medication, dose count **Anticoagulant medication**, dose count **Anti-infective medication**, dose count **Gastrointestinal medication,** dose count **Cardiac medication**, dose count **Anti-hyperglycemic medication**, dose count **Hematocrit, %;** mean, high, low **Hemoglobin, g/dL**; mean, high, low **Platelet count, 10^3cells/*u*L**; mean, high, low **Bicarbonate, mmol/L**; mean, high, low **Creatinine, mg/dL**; mean, high, low **Potassium, mmol/L**; mean, high, low **Magnesium, mmol/L**; mean **Pro-thrombin time, seconds**; mean **Partial thromboplastin time, seconds**; mean	Stage 2	Low frequency k= 27
Heart rate, beats/min **Respiratory rate, breaths/min** **Percent oxygen saturation, %** **Blood pressure, mmHg; systolic, diastolic, mean arterial**	Stage 3	High frequency k=36

3.4.4 Machine Learning Plan

Our machine learning process plan (Figure 15) involved periods of **observation** during SDU stay (during which potentially predictive LF and HF data were extracted) and **prediction** (specified time windows preceding occurrence of HA-VTE). We used the dynamic and static data in 24-hour windows, at various lead times ($PTW_{48\text{-}24}$ and $PTW_{24\text{-}0}$) before T_0. The models started with Stage 1 static data and then progressed to Stage 2 (adding the LF data dynamically accrued in the time window) and the Stage 3 (adding the HF data dynamically accruing in the time window), as shown in Figure 14. These 3 stages of modeling were applied to each of the two prediction time windows before T_0 ($PTW_{48\text{-}24}$ and $PTW_{24\text{-}0}$) and for each clinical rationale cohort (S +/- and C+/-). Our objective was to build the best attainable model for predicting the HA-VTE outcome event for each time window, and for each clinical rationale cohort.

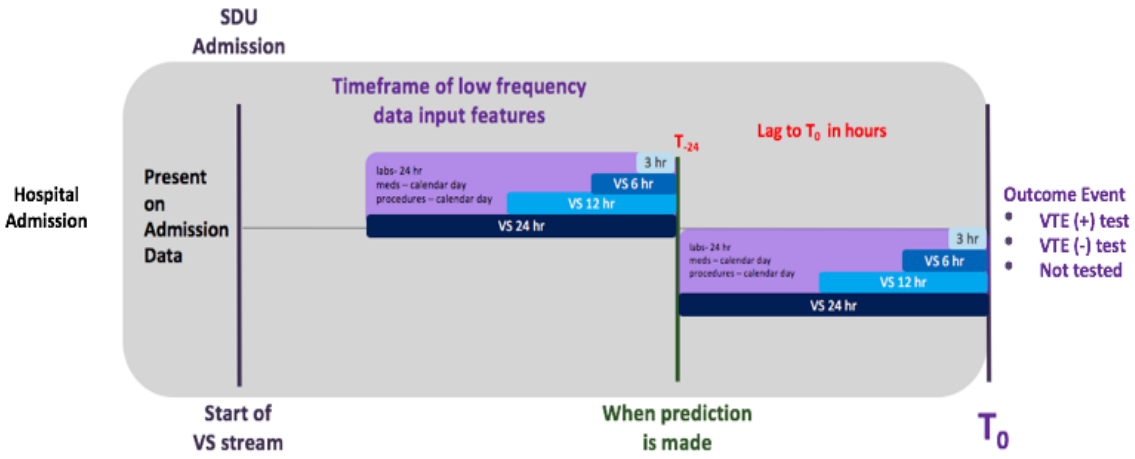


Figure 15. Review of Machine Learning Process

3.4.4.1 Supervised ML Algorithm and Feature Selection

Data set size and structure and important factors in ML approaches and no single algorithm works best for every problem. This is especially true with supervised learning and the rationale

behind our approach to try several different algorithms for the clinical questions proposed. Supervised machine learning candidate algorithm selection was based on data characteristics and constraints (data missingness, level of measurement, class imbalance) and a desire for model interpretability. In ML, bias and variance provide the tools to understand the behavior of ML algorithms in pursuit of predictive performance. Bias are the simplifying assumptions made by a model to make the target function easier to learn. Bias can aid in generalization. Variance is a type of error that occurs due a model's sensitivity to small fluctuations in the training set. High variance can result in overfitting which impacts generalizability and performance of the model when applied to another population or set of data. The bias-variance trade-offs of different classification algorithms were carefully considered, and the ones ultimately selected for use are detailed below.

Binary logistic regression (BLR) is a simple classifier that can provide a good baseline model for comparison. Interpretability of BLR is an advantage, however this approach is likely to produce a model with higher bias (underfitting). LASSO (least absolute shrinkage and selection operator) L1 regularization was used for feature selection and to minimize overfitting. Because it is still a linear modeling approach, a weakness of BLR is that it can underperform when there are non-linear decision boundaries and can fail to capture more complex data relationships.

Naïve Bayes (NB) is another highly interpretable classifier that has been shown to perform well on small amounts of training data, to be relatively robust to missing data, and it has few tunable parameters. Similar to BLR, NB tends to build models with higher bias, but low variance. A potential weakness of NB is that it makes the assumption of conditional independence between its features. For this modeling approach, we fit the model using all predictor features without using any feature selection approaches. BLR and NB are higher bias, lower variance algorithms.

Random Forest (RF) is an ensemble method that uses bagging (bootstrap aggregating) to combine many decision trees in parallel enabling its ability to produce models with low bias and moderate variance.

Gradient Boosted Decision Trees (GBDT) is another ensemble method that uses boosting, combining a series of sequentially connected decision trees, each learning from the errors of the previous one, to produce a highly efficient and accurate model.

The strengths and weaknesses of these two ensemble decision tree algorithms are similar. Powerful classifiers fairly robust to outliers and missing data, RF and GBDT algorithms handle categorical and continuous data well and are interpretable, albeit less so than BLR and NB. Iterative modeling with new data can be more challenging for these algorithms and they require more hyperparameter tuning to refine model performance. Both include a feature importance estimate capability for built in feature selection. RF and GBDT algorithms are characterized by lower bias and moderate variance.

3.4.5 Model Construction

Machine learning application analyses were run using R-version 4.0.2 (R Core Team, 2020) and Python (Python Software Foundation, https://www.python.org, Sklearn library) programming language.

3.4.5.1 Models for Clinical Rational #1

Dataset. The sample cohort including patients "suspected (tested for HA-VTE)" and "not suspected (not tested for HA-VTE)" by clinicians comprised the dataset for this set of S +/- models (n= 2370). Prevalence of tested cases was 32%.

Stage-wise Model Construction

Stage 1. The ML algorithms and the feature selection methods previously outlined were applied to the training data static (Stage 1) variables only. The performance of each trained model was tested via stratified 10-fold cross validation. The average model performance across the 10 validation tests is reported for each model, BLR, NB, RF, GBDT. Stage 1 analysis resulted in development of a total of four models to classify overall risk of clinician suspicion at time of hospital admission.

Stage 2. Next, LF variables extracted from $PTW_{48\text{-}24}$ and $PTW_{24\text{-}0}$were added to the static data, producing a data set for Stage 2 $PTW_{48\text{-}24}$ models (Static data + $PTW_{48\text{-}24}$ LF data) and for Stage 2 $PTW_{24\text{-}0}$ models (Static data + $PTW_{24\text{-}0}$ LF data). The same ML algorithms and feature selection methods used in Stage 1 were applied to these Stage 2 data and then the performance of each trained model was tested and reported using the same methods outlined for Stage1 above. Stage 2 analysis resulted in development of a total of eight models to classify risk of clinician suspicion using static and LF data: four models (BLR, NB, RF, GBDT) 24 hours in advance of HA-VTE event outcome ($PTW_{48\text{-}24}$) and four at time of HA-VTE event outcome ($PTW_{24\text{-}0}$).

Stage 3. In the last stage, HF variables extracted from $PTW_{48\text{-}24}$ and $PTW_{24\text{-}0}$ were added to the Stage 2 data, producing a data set for Stage 3 $PTW_{48\text{-}24}$ models (Static data + $PTW_{48\text{-}24}$ LF data + $PTW_{48\text{-}24}$ HF data) and for Stage 3 PTW-0 models (Static data + $PTW_{24\text{-}0}$ LF data + $PTW_{24\text{-}0}$ HF data). The same ML algorithms and feature selection methods were applied to these Stage 3 data and then the performance of each trained model was tested and reported using the same methods used in the previous two stages. Stage 3 analysis resulted in development of a total of eight models to classify risk of clinician suspicion using static, LF, and HF data: four models (BLR, NB, RF, GBDT) 24 hours in advance of test ($PTW_{48\text{-}24}$) and four at time of test ($PTW_{24\text{-}0}$).

The sample size for the $PTW_{48\text{-}24}$ models was smaller than the $PTW_{24\text{-}0}$ sample as it does not include patients admitted to the SDU <24 hours before the time of their VTE outcome event.

3.4.5.2 Models for Clinical Rational #2

Dataset. The sample cohort including patients with "confirmed disease (radiologic test results positive for HA-VTE)" and "confirmed no disease (radiologic test results negative for HA-VTE)" comprised the dataset for this set of C+/- models (n= 754). Class imbalance was a reality of this dataset, with the outcome variable of interest (HA-VTE positive diagnostic test result) being in the minority class (6.2% prevalence). This known constraint was a consideration during study design that strongly influenced algorithm selection choices. Synthetic minority oversampling technique (SMOTE)[116] was employed to create a more balanced cohort sample for training, and a variety of balance ratios were explored. The balanced sample with a ratio of 49% positive test cases, 51% negative test cases was used for model training. The performance of each trained model was evaluated via 10-fold stratified cross validation using the original sample cohort.

Stage-wise Model Construction

Using the aforementioned training and testing sets, the C+/- models were developed using the same stage-wise process (Stage 1, 2, 3) as described for the S+/- models in section 3.4.5.1

3.4.6 Model Assessment

Model capability to predict the specified HA-VTE event outcome for each model set (S+/- and C+/-) was assessed by a variety of metrics. **Recall** (true positive rate [TPR]/sensitivity), **specificity** (true negative rate [TNR]), **precision** (positive predictive value [PPV]), **F1 score** (harmonic mean of the precision and recall), **area under the receiver operating characteristic**

(AUROC) curve and **area under the precision recall curve** (AUPRC) were reported for each model. This study required evaluation of the following 40 models:

24 Models for $PTW_{24\text{-}0}$ (between 24-0 hours before T_0 or T_0 models)

- **12 T_0 models for the S+/- cohort:**
  - 4 models (BLR, NB, RF, GBDT) with Stage 1 static data
  - 4 models (BLR, NB, RF, GBDT) with Stage 2 LF data
  - 4 models (BLR, NB, RF, GBDT) with Stage 3 HF data
- **12 T_0 models for the C+/- cohort:**
  - 4 models (BLR, NB, RF, GBDT) with Stage 1 static data
  - 4 models (BLR, NB, RF, GBDT) with Stage 2 LF data
  - 4 models (BLR, NB, RF, GBDT) with Stage 3 HF data

16 Models for $PTW_{48\text{-}24}$ (between 48-24 hours before T_0 or T_{-24} models)

- **8 T_{-24} models for the S+/- cohort:**
  - 4 models (BLR, NB, RF, GBDT) with Stage 2 LF data
  - 4 models (BLR, NB, RF, GBDT) with Stage 3 HF data
- **8 T_{-24} models for the C+/- cohort:**
  - 4 models (BLR, NB, RF, GBDT) with Stage 2 LF data
  - 4 models (BLR, NB, RF, GBDT) with Stage 3 HF data

3.4.6.1 A Priori Metrics

The decision was made, a priori, to identify the best performing model classifier in each stage by comparing the aforementioned metrics. Model performance was evaluated through the lens of the clinical gap we aimed to address: HA-VTE disease pathology, a leading cause of

unplanned hospital death, evolves slowly over the course of hospitalization, is associated with symptoms common to many conditions present in hospitalized patients, and when not identified early, can lead to death. Identification is recognized as a clinical challenge and yet no current VTE risk models incorporate the progressive accrual of dynamic patient data and pattern evolution over time in their modeling approaches for risk prediction.

HA-VTE is a diagnosis associated with increased morbidity, mortality, hospital length of stay, and health care costs, thus the cost of missing a positive diagnosis is high. Accurate model identification of positive cases is critically important. We are willing to tolerate some false positives, because we cannot allow false negatives (true cases that are not identified as cases and thus missed by clinicians). Cases classified as false negatives by predictive models equate to failure to rescue in the clinical setting.

Prevalence of the HA-VTE outcome event in our two sample cohorts is reflective of population prevalence, a strength of our approach, and another factor influencing model performance evaluation. In each clinical rationale cohort/model set, we have imbalanced data, with the cases we seek to predict being in the minority class. With awareness of these issues, we prioritized metrics immune to class imbalance. The following metrics were prioritized in ranking model performance for each prediction development stage.

The sensitivity (recall/TPR) metric indicates how well the model is able to predict cases of interest out of all the actual cases and the specificity (TNR) metric indicates how well the model is able to identify controls, out of all actual controls. Among the patients the model identifies to be cases, the precision metric (PPV) indicates how many are truly cases. As these two metrics exist in a trade-off, (increasing one parameter leads to decreasing the other), the harmonic mean of the two metrics, the F1 score, summarizes the models' precision-recall performance in a single metric.

AUPRC was also examined. A model achieves perfect AUPRC (1.0) when able to identify all cases (perfect recall), without including any false positives (perfect precision). The AUPRC score must be evaluated in comparison to the AUPRC baseline, which is equal to the fraction of true cases (0.36 for the S+/- models and 0.064 for the C+/- models) in the sample. Finally, area under the receiver operating characteristic (AUROC) was examined for tradeoffs in model sensitivity and specificity.

Model evaluation informs model optimization (a component of model development) as models should be optimized and parameters tuned based on the planned utility and application. Fine tuning and optimization of these quantity of models exceeds the scope of the study, however, these findings will inform future work on this issue.

3.5 Results

3.5.1 Descriptive Statistics

Our total sample for model development included 2370 adult SDU patients. Among the three HA-VTE outcome event classes (Table 6), there were no significant differences in age, co-morbidity burden, or overall hospital LOS. Notable differences between the groups were that tested patients were more likely to be admitted as a trauma, and to have higher mean heart and respiratory rates and lower blood pressure and oxygen saturation levels.

Table 9. Sample Characteristics: Total and by Hospital Acquired Venous Thromboembolism Event Outcome

	Total sample	Tested Patients		Not Tested	
Characteristic		**Tested HA-VTE (+)**	**Tested HA-VTE (-)**		
	n = 2370	n = 47	n = 713	n = 1614	*p* value
Hospital admission LOS in days, *mean (SD)*	12.29 (14.10)	15.0 (9.17)	13.05 (16.76)	11.88 12.86)	.074
Time to test in hours, *mean (SD)*	62.97 (73.72)	67.15 (53.72)	63.90 (77.19)	62.4 (72.67)	.839
Age in years, *mean (SD)*	59.52 (19.83)	56.49 (17.89)	59.17 (19.83)	59.76 (19.89)	.461
Charlson Comorbidity Index Deyo method (CCI-D), *mean (SD)*	1.27 (1.69)	0.96 (1.33)	1.27 (1.61)	1.27 (1.739)	.448
Race, *n* (%)					
White	1719 (72)	38 (80.9)	527 (73.9)	1154 (71.7)	.234
Black	283 (11.9)	4 (8.5)	72 (10.1)	201 (12.9)	.244
Mean Vital Sign Data 24 hr. in advance of T_{-24}	**Total Sample**	**Tested Patients**		**Not Tested**	
Heart rate, *mean (SD)*	85.11 (14.65)		88 (16.13)	83.84 (13.76)	**.000***
Respiratory rate, *mean (SD)*	18.73 (3.60)		19.12 (3.81)	18.56 (3.49)	**.028***
SpO_2, *mean (SD)*	96.15 (2.29)		95.95 (2.41)	96.24 (2.22)	.094
Diastolic blood pressure, *mean (SD)*	67.53 (11.76)		67.10 (11.79)	67.72 (11.75)	**.029***
Systolic blood pressure, *mean (SD)*	128.82 (17.18)		127.43 (17.31)	129.42 (17.09)	.077
Mean Vital Sign Data 24 hr. in advance of T_0	**Total Sample**	**Tested Patients**		**Not Tested**	
Heart rate, *mean (SD)*	84.48 (14.58)		87.43 (15.79)	83.05 (13.73)	**.000***
Respiratory rate, *mean (SD)*	18.74 (3.61)		19.31 (3.84)	18.46 (3.46)	**.000***
SpO_2, *mean (SD)*	96.08 (2.45)		95.74 (2.40)	96.25 (2.5)	**.000***
Diastolic blood pressure, *mean (SD)*	68.42 (11.90)		67.59 (10.97)	68.82 (12.33)	**.015***
Diastolic blood pressure, *mean (SD)*	129.17(17.57)		126.59 (17.12)	130.41 (17.65)	**.000***
Prevalence of Present-on-Admission Diagnoses, *n* (%)	**Total Sample**	**Tested Patients**		**Not Tested**	
Trauma admission	1017 (42.9)		355 (46.8)	662 (41.0)	**.002[a]***
Hypertension	913 (38.5)		293 (38.6)	620 (38.4)	.985[a]
Ischemic heart disease	465 (19.6)		133 (17.5)	332 (20.6)	.138[a]
Diabetes Mellitus (DM)	432 (18.2)		143 (18.8)	289 (17.9)	.790[a]
Chronic pulmonary disease	419 (17.7)		137 (18.1)	282 (17.5)	.574[a]
Septicemia	264 (11.1)		83 (11.0)	181 (11.2)	.865[a]
Congestive heart failure (CHF)	218 (9.2)		77 (10.1)	141 (8.7)	.520[a]
Prior myocardial infarction (MI)	189 (8.0)		56 (7.4)	133 (8.2)	.549[a]
Lower Extremity Fracture	179 (7.6)		69 (9.1)	110 (6.8)	**.047[a]***
Malignancy	104 (4.4)		45 (6.0)	59 (3.7)	**.011[a]***
Obesity	100 (4.2)		38 (5.0)	62 (3.8)	.181[a]
DM with complications	60 (2.5)		27 (3.6)	33 (2.0)	**.027[a]***
Irritable Bowel Syndrome (IBS)	57 (2.4)		20 (2.8)	37 (2.3)	.601[a]
Coagulation disorder	58 (2.4)		14 (1.9)	44 (2.7)	.199[a]
Chronic venous Insufficiency	51 (2.2)		13 (1.8)	38 (2.4)	.430[a]
Prior stroke	46 (1.9)		13 (1.8)	33 (2.0)	.587[a]
Acute MI	38 (1.6)		12 (1.6)	26 (1.6)	.966[a]
Spinal Cord Injury	26 (1.1)		10 91.4)	16 (1.0)	.470[a]
Stroke	24 (1.0)		4 (0.6)	20 (1.2)	.126[b]
Varicose veins	5 (0.2)		0	5 (0.3)	.146[b]
HIV	2 (0.1)		1 (0.1)	1 (0.1)	.536[b]

p-values represent ANOVA comparison between Tested HA-VTE (+), Tested HA-VTE (-), and Not tested unless otherwise noted

[a] Pearson chi-square, [b] Fisher's exact test, * $p < .05$

3.5.2 Model Performance

3.5.2.1 Clinical Rationale #1 Models

Patients Clinicians Suspect of HA-VTE and Test vs Not Suspected (and not tested)

We compared four ML algorithm classifiers (BLR, RF, NB, GBDT) using 65 features, (Table 5), building models using a stage-wise approach to enable evaluation of model performance based on degree of granularity of dynamic clinical data variables. Table 7 shows the average of 10-fold cross validation performance results, reported at the 0.5 cut point, for the models developed based on clinical rationale #1, discriminating patients who clinicians suspected of HA-VTE and tested from patients they did not suspect and did not test at time of event (T_0) and 24 hours in advance of the event (T_{-24}).

Comparison of Performance Evaluation Metrics

Model performance varies by algorithm. The trend across all clinical rationale #1 (S+/-) models, at both prediction time points, was an improvement in sensitivity, specificity, precision, F1-score, AUPRC, and AUROC performance with the progressive addition of dynamic clinical data.

(S+/-) Model Performance at T_0, Time of HA-VTE Outcome Event

At time of hospital admission, using only static data, sensitivity performance for all models was no better than random guessing. As more dynamic data was added, the ability to identify positive cases (those suspected of HA-VTE risk and tested) improved. Sensitivity of the NB model improved from 18% in Stage 1 using only static data, to 73% with the addition of dynamic LF data, to 77% when HF dynamic VS data is added. At the T_0 prediction time point using data in the PTW_{24-0}, the sensitivity of the three other models followed the same trend of improvement with each stage, but with less impressive TPR identification ability (BLR 65%, RF 46%, GBDT 49%).

With an AUPRC of 0.56, a PPV of 59%, and an AUROC of 0.66, NB using static and dynamic LF and HF data, was identified as the most sensitive model at this prediction time point.

With regard to specificity, NB (75%), RF (63%) and GBDT (74%) had notable ability to identify patients at low risk for HA-VTE. However, as more dynamic data was added, specificity performance of the NB model was traded for improved sensitivity, as noted above. BLR specificity improved from 50% (Stage 1) to 68% with the addition of HF dynamic data in Stage 3. The RF and GBDT ensemble decision tree models performed the best, with specificities of 78% (RF Stage 2), 81% (RF Stage 3) and 76% (GBDT Stage 2), 79% (GBDT Stage 3) respectively, as more dynamic features were added. With an AUPRC of 0.60, a PPV of 62%, and an AUROC of 0.70, RF using static and dynamic LF and HF data, was identified as the most specific model at this prediction time point.

(S+/-) Model Performance at T_{-24}, 24 hours in Advance of HA-VTE Outcome Event

The NB model, with a sensitivity of 72% using static and LF data and 76% with the addition of HF data, (AUPRC .52, PPV 46%, AUROC .60) had the best ability to identify true positive cases (those suspected of HA-VTE risk and tested) 24 hours in advance of the diagnostic test being conducted (T_{24} with the $PTW_{48\text{-}24}$). BLR model sensitivity was 55% using static and LF data and increased to 59% with the addition of HF data in stage 3. Sensitivity performance of RF and GBDT at this prediction time was unimpressive.

With very similar performance metrics, the most specific S+/- models at the T_{-24} prediction time were the ensemble decision trees. RF and GBDT demonstrated 75% and 72% sensitivity, respectively when dynamic LF data was added. The addition of dynamic HF VS data in Stage 3 improved ability to identify true negatives to 80% for both models. Neither NB nor BLR models

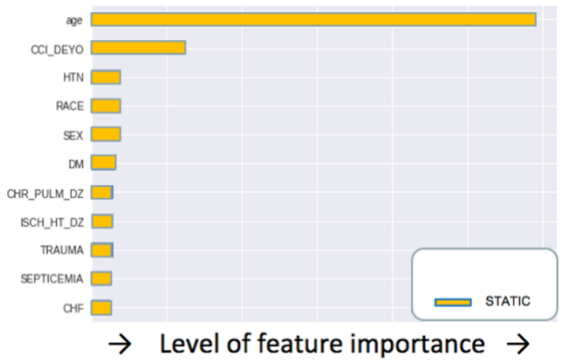
were able to identify true negative cases with high levels of confidence, (47% and 54%, respectively), 24 hours in advance of HA-VTE outcome.

Table 10. Model Performance for Clinical Rationale #1: Classifying Patients Who Clinicians Suspect of Hospital Acquired Venous Thromboembolism (HA-VTE) and Test, from Patients They Do Not Suspect

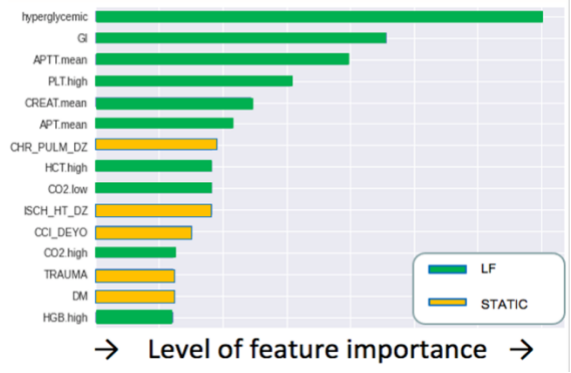
	Data Granularity Progression				
	Hospital Admission	24 hr before event T_{-24}		At event occurrence T_0	
	Static	T_{-24} Static + LF data	T_{-24} Static + LF +HF	T_0 Static + LF data	T_0 Static + LF+ HF
Binary logistic regression					
AUPRC	0.37	0.48	0.50	0.52	0.57
Precision (PPV)	0.39	0.44	0.47	0.50	0.54
Recall (TPR/sensitivity)	**0.47**	**0.55**	**0.59**	**0.61**	**0.65**
F1-score	0.42	0.48	0.52	0.55	0.59
Specificity (TN)	**0.50**	**0.50**	**0.54**	**0.60**	**0.63**
AUROC	0.47	0.60	0.60	0.66	0.68
Naïve Bayes					
AUPRC	0.36	0.50	0.52	0.52	0.56
Precision (PPV)	0.31	0.46	0.46	0.46	0.59
Recall (TPR/sensitivity)	**0.18**	**0.72**	**0.76**	**0.73**	**0.77**
F1-score	0.23	0.57	0.58	0.58	0.59
Specificity (TN)	**0.75**	**0.40**	**0.47**	**0.38**	**0.50**
AUROC	0.47	0.60	0.60	0.63	0.66
Random Forest					
AUPRC	0.50	0.57	0.60	0.61	0.60
Precision (PPV)	0.47	0.54	0.60	0.59	0.62
Recall (TPR/sensitivity)	**0.50**	**0.40**	**0.42**	**0.46**	**0.46**
F1-score	0.48	0.46	0.50	0.52	0.53
Specificity (TN)	**0.63**	**0.75**	**0.80**	**0.78**	**0.81**
AUROC	0.59	0.66	0.66	0.69	0.70
Gradient Boosted Decision Trees					
AUPRC	0.39	0.53	0.60	0.54	0.61
Precision (PPV)	0.61	0.52	0.60	0.55	0.62
Recall (TPR/sensitivity)	**0.20**	**0.42**	**0.43**	**0.44**	**0.49**
F1-score	0.25	0.46	0.50	0.49	0.55
Specificity (TN)	**0.74**	**0.72**	**0.80**	**0.76**	**0.79**
AUROC	0.51	0.64	0.65	0.67	0.71
	Time of hospital admission	24 hours before HA-VTE event		HA-VTE event occurrence	

Model Feature Importance

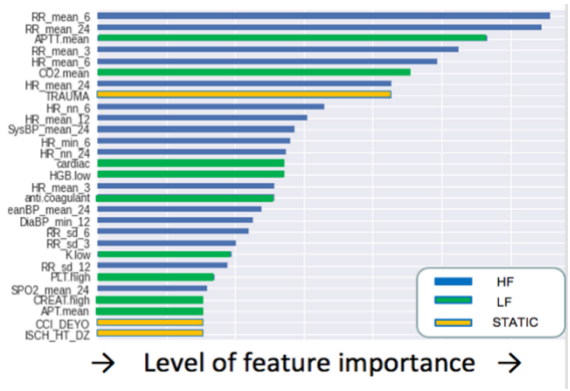
As shown in Figure 16, at time of admission (Figure Panel 1), age and comorbidity score (CCI-D) were the features identified by the models most important to classification performance. As dynamic medication and lab data was added in Stage 2 (Figure Panel 2), age was no longer identified as an important feature. Antihyperglycemic and gastrointestinal prophylaxis medications and laboratory values, specifically thrombin and prothrombin platelets and hematocrit levels, were the clinical features identified to be of greatest importance in this stage. With the addition of HF VS data (Figure Panel 3), RR and HR features (mean, monitoring frequency, and variability) dominated in importance. Notably, trauma admission was a feature identified as important by all models in all three stages.



Panel 1
Stage 1: Static Features
Prediction time: T_0
Model: Random Forest



Panel 2
Stage 2: Static + Low Frequency
Prediction time: T_{-24}
Model: Naïve Bayes



Panel 3
Stage 3: Static + Low + High Frequency
Prediction time: T_{-24}
Model: Naïve Bayes

Figure 16. Feature Importance Performance for Clinical Rationale #1: S+/- Models

3.5.2.2 Clinical Rationale #2 Models

Confirmed Positive HA-VTE Diagnosis vs Confirmed Negative for HA-VTE (C+/-)

Table 8 shows the average of 10-fold cross validation performance results, reported at the 0.5 cut point, for the models developed based on clinical rationale #2, discriminating patients with a confirmed positive HA-VTE diagnostic test result from those with a confirmed negative result at time of event (T_0 and PTW $_{48\text{-}24}$) and 24 hours in advance of the event (T_{-24} and PTW $_{48\text{-}24}$).

Comparison of Performance Evaluation Metrics

Model performance varies by algorithm, but the trend across all clinical rationale #2 (C+/-) models, at both prediction time points, was an improvement in sensitivity, specificity, precision, F1-score, AUPRC, and AUROC performance with the progressive addition of dynamic clinical data.

Model Performance at T_0, Time of HA-VTE Outcome Event

At time of hospital admission using only static data (Stage 1), the ensemble decision tree models, RF (sensitivity 81%) and GBDT (sensitivity 80%), performed similarly with regard to their ability to accurately predict true positive diagnoses at T_0. As more dynamic data was added, these sensitivity metrics improved for RF to 86% at Stage 2 and 89% (Stage 3) and for GBDT to 89% (Stage 2) and 93% (Stage 3). Naïve Bayes model at time of admission had a sensitivity of only 56% at Stage 1, but sensitivity performance improved notably with the addition of dynamic data to 93% at Stage 2 and 94% at Stage 3. Sensitivity of the BLR model was only 29% with only static data at Stage 1 and improved to 66% at Stage 2 and 94% at Stage 3 with the addition of dynamic data. With an AUPRC of 0.81, a PPV of 84%, and an AUROC of 0.90, BLR, using static and dynamic LF and HF data, was identified as the most sensitive model at this prediction time point.

Regarding specificity, RF and GBDT models performed the best with specificities of 94% (Stage 1 RF), 96% (Stage 2 RF), 97% (Stage 3 RF) and 95% (Stage 1 GBDT), 92% (Stage 2 GBDT), 95% (Stage 3 GBDT). BLR specificity improved from 54% (Stage 1) to 87% with the addition of HF dynamic data in Stage 3. NB, specificity was 65% in Stage 1, but while the addition of dynamic data improved sensitivity (as noted above), specificity decreased to 39% in Stage 3. With an AUPRC of 0.97, a PPV of 96%, and an AUROC of 0.93, RF, was identified as the most specific model at this prediction time point.

Model Performance at T_{-24}, 24 hours in Advance of HA-VTE Outcome Event

The NB model, with a sensitivity of 88% using static and LF data and 91% with the addition of HF data, had the best ability to identify true positive HA-VTE cases 24 hours in advance of the diagnostic test being conducted (T_{-24}, PTW_{48-24}). RF and GBDT demonstrated a sensitivity of 74% and 82% respectively at Stage 2, but performance for both dropped to 70% and 73%with the addition of Stage 3 HF data, raising the suspicion of overfitting. BLR model sensitivity was 44% using Stage 2 static and LF data and increased to 62% with the addition of HF data in stage 3.

The most specific models at the T_{-24} prediction time were the ensemble decision trees. RF and GBDT demonstrated 94% and 95% specificity, respectively, at Stage 2, but performance for both dropped to 90% with the addition of HF data at Stage 3, raising the suspicion of overfitting. With regard to specificity, when HF Stage 3 data was employed, neither NB nor BLR performed well (34% and 66%, respectively).

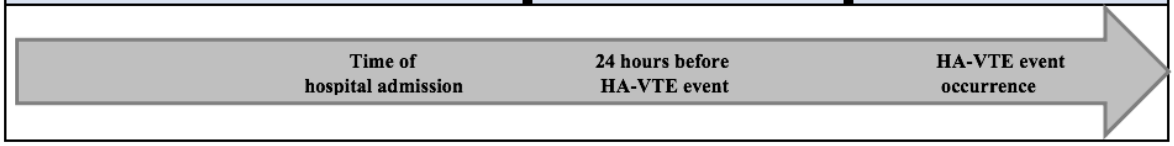
Model Feature Importance

As shown in Figure 17, at time of admission (Figure Panel 1), age and comorbidity score (CCI-D) were the features identified by the models most important to classification performance. As dynamic medication and lab data was added in Stage 2 (Figure Panel 2), age was no longer

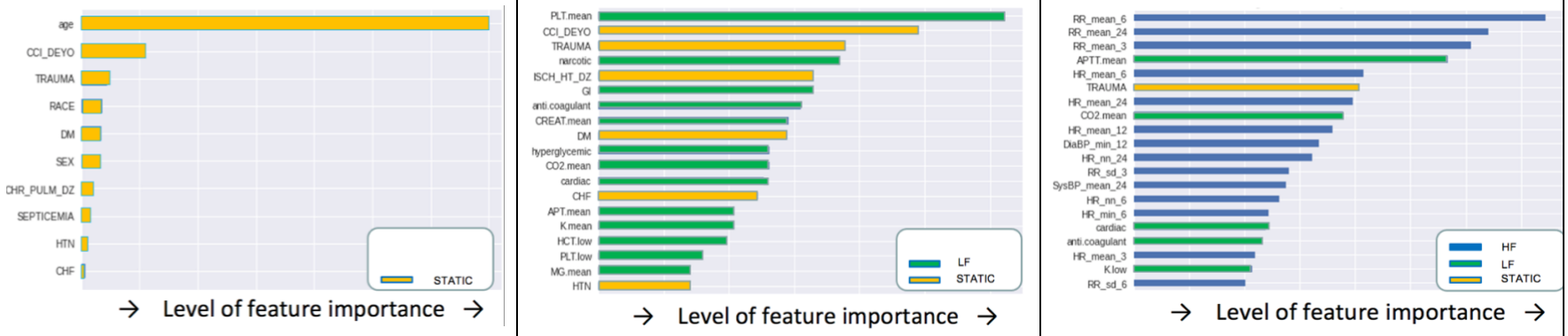
identified as an important feature and narcotic and anticoagulant medication counts, mean platelet, age, and CCI-D scores were the clinical features identified to be of greatest importance in this stage. With the addition of HF VS data (Figure Panel 3), RR and HR features (mean, monitoring frequency, and variability) dominated in importance. As with the S+/- models, trauma admission was a feature identified as important by all models in all three stages.

Table 11. Model Performance for Clinical Rationale #2: Classifying Patients with a Confirmed Positive HA-VTE Diagnosis from Patients with a Confirmed Negative Test

	Data Granularity Progression				
	Hospital Admission	24 hr before event T_{-24}		At event occurrence T_0	
	Static	T_{-24} Static + LF data	T_{-24} Static + LF +HF	T_0 Static + LF data	T_0 Static + LF+ HF
Binary logistic regression					*
AUPRC	0.51	0.67	0.60	0.69	0.81
Precision (PPV)	0.48	0.55	0.56	0.69	0.84
Recall (TPR/sensitivity)	**0.29**	**0.44**	**0.62**	**0.66**	**0.94**
F1-score	0.37	0.49	0.58	0.67	0.89
Specificity (TN)	0.79	0.75	0.66	0.79	0.87
AUROC	0.54	0.59	0.64	0.72	0.90
Naïve Bayes					
AUPRC	0.63	0.67	0.67	0.68	0.77
Precision (PPV)	0.53	0.46	0.53	0.49	0.53
Recall (TPR/sensitivity)	**0.56**	**0.88**	**0.91**	**0.93**	**0.94**
F1-score	0.54	0.60	0.65	0.65	0.68
Specificity (TN)	0.65	0.27	0.34	0.33	0.39
AUROC	0.60	0.58	0.62	0.63	0.68
Random Forest					
AUPRC	0.91	0.90	0.94	0.97	0.97
Precision (PPV)	0.90	0.90	0.83	0.94	0.96
Recall (TPR/sensitivity)	**0.81**	**0.74**	**0.70**	**0.86**	**0.89**
F1-score	0.85	0.81	0.77	0.88	0.92
Specificity (TN)	0.94	0.94	0.90	0.96	0.97
AUROC	0.87	0.74	0.81	0.90	0.93
Gradient Boosted Decision Trees					
AUPRC	0.89	0.90	0.94	0.97	0.98
Precision (PPV)	0.92	0.80	0.87	0.89	0.93
Recall (TPR/sensitivity)	**0.80**	**0.82**	**0.73**	**0.89**	**0.93**
F1-score	0.86	0.76	0.78	0.89	0.92
Specificity (TN)	0.95	0.95	0.90	0.92	0.95
AUROC	0.88	0.76	0.81	0.90	0.94



Average of 10-fold cross validation results reported at 0.5 cut point
AUPRC area under the precision recall curve, baseline reference 0.064
AUROC area under the receiver operating characteristic curve



Panel 1	Panel 2	Panel 3
Stage 1: Static Features Prediction time: T_0 Model: Random Forest	Stage 2: Static + Low Frequency Prediction time: T_{-24} Model: Naïve Bayes	Stage 3: Static + Low + High Frequency Prediction time: T_{-24} Model: Naïve Bayes

Figure 17. Feature Importance Performance for Clinical Rationale #2: C+/- Models

3.6 Discussion

This hypothesis driven study is the first to our knowledge to incorporate the progressive addition of dynamic clinical data routinely accrued during hospitalization in the development of models to predict risk of new onset, HA-VTE. No published research studies have explored clinical data features predictive of new-onset, HA-VTE in any time frame intervals. One study by Posch et al., identified dynamic patterns in monthly d-dimer lab test values predictive of cancer associated VTE in the outpatient oncology population. [111] These findings were independent of other established VTE risk factors indicating the value and utility of using dynamic, longitudinal individual clinical data to inform risk prediction for this challenging disease pathology.

Additionally, it is the first to endeavor to predict risk within specific time windows in advance of HA-VTE diagnostic test conduction and results. Our foundational models provide emerging evidence that the addition of dynamic clinical data improves HA-VTE risk prediction, that HA-VTE risk stratification can be accomplished within specific time intervals, and insights on clinical feature importance associated with HA-VTE outcome over time.

HA-VTE is a disease pathology known to have low incidence rates of confirmed disease in hospitalized patients, yet patients who develop this complication are at risk of death, high recurrence rates, long term complications, and increased medical cost burden.[124] For these reasons, algorithms with the ability to handle different types of clinical data and capable of performing well on small amounts of training data were chosen. Strategies for approaching model building with imbalanced data (reflective of the real-world clinical scenarios) and for evaluating model performance based on both data constraints and the high cost of missing this complication

were employed. For both clinical rationales (S+/-, C+/-) and at both prediction time points (T_{-24} and T_0), all models demonstrated that the addition of dynamic high frequency VS data improves ability to improve prediction for risk of the complication of VTE during SDU admission. Precision-recall ability in models with dynamic (high and low frequency) clinical data was greatly improved results relative to models with static data only. This ability to discriminate between groups is especially notable in this sample of SDU patients who were fairly homogenous with respect to age, level of acuity and reasons for hospital admission.

3.6.1 Comparison of Study Findings to Current VTE RAM's

The Caprini model, developed in 1991, and validated on general surgical, medical, and plastic surgery populations, is one of the most widely used VTE RAMs. When all 32-items required by the tool are completed, including several specialized coagulopathy and genetic serum laboratory tests, sensitivities as high as 88.6% have been reported in retrospective validation studies. [125, 126] However, the specificity of this model is reported to be only in the range of 20%. Recommended by the CHEST 2012 VTE prevention guidelines, [68] the Caprini RAM has been criticized for being cumbersome to implement clinically and for the complexity of data clinicians are required to collect. Despite its wide endorsement and use in clinical practice for the past three decades, prospective studies demonstrating reports of significant reduction in HA-VTE rates as a result of Caprini model use are few. [126, 127]. The best reported sensitivity of the other commonly use VTE RAMs are 75% for the Geneva model, 61.8%for the Padua model and 63.3% for the IMPROVE-DD model, but these are validated for use in medical populations only.[119, 128, 129] In comparison, using only data that is accrued in routine daily clinical care, and at a time point 24 hours in advance of a HA-VTE diagnostic test being ordered by a clinician and conducted, our

S+/- NB model demonstrated the ability identify suspected general medical-surgical patients with a sensitivity of 76%. At the same time point, for general medical-surgical patients already identified by clinicians as being at increased risk, our C+/- NB model demonstrated the ability to identify patients with a confirmed positive HA-VTE diagnosis with a sensitivity of 91%. These findings, based on commonly collected data, predicting outcome risk 24 hours in advance of the event, hold promise that these models could, with further development and testing, provide clinicians with personalized information that could support earlier intervention and rescue.

Other VTE RAM's with greater *specificity* performance metrics, such as the IMPROVE-DD and Kucher RAM's, with best reported specificities of 70.7% and 85.7% respectively, can be useful in risk stratification for ruling out patients at low risk of HA-VTE. Identification of low-risk patients can spare those individuals from unnecessary risks associated with anticoagulant prophylaxis use and can better inform the utilization of scare clinical resources such as continuous monitoring and acuity-based nurse staffing ratios. The Kucher model was developed to be an alert tool, and when integrated into the Brigham and Women Hospital's (BWH) electronic medical record it was successful in decreasing their local HA-VTE rates by 41%.[129] However, in external validation studies, the Kucher model has failed to perform as well, raising questions about its generalizability to populations with lower rates of malignancy.[119] A rigorous validation study lead by Greene et al., using a cohort of over 60,000 medical patients from multiple institutions, evaluated the performance ability of the IMPROVE-DD, Kucher, Padua, and Geneva RAMs and found the performance ability of these models to be suboptimal, with limited ability to identify patients at-risk for VA-VTE.[126] Our models use a novel approach that incorporates routinely collected data accruing dynamically during hospitalization and follows the HA-VTE evolution timeline. At a time point 24 hours in advance of a HA-VTE diagnostic test being ordered by a

clinician and conducted, our S+/- cohort RF model (patients identified by clinicians as not suspected of HA-VTE and thus not tested) demonstrated the ability identify true negative cases in general medical-surgical patients with a specificity of 80%. At the same time point, among general medical-surgical patients already identified by clinicians as being at increased risk, our C+/- cohort RF and GBDT models demonstrated the ability to identify patients with a confirmed negative HA-VTE diagnosis with a specificity of 90%.

3.6.2 Identification of Features Important to Model Performance

Mean vital signs values for HR, RR, and BP were significantly different between tested and not-tested patients in the time period 48-24 hours before the HA-VTE event outcome. And closer to the outcome, mean vital signs accrued during the 24-0 hour time period were significantly different between the two groups for measurements (HR, RR, SpO_2, BP). This descriptive data different was also seen in the feature importance patterns identified by the models.

When we added to the data used for modeling, features of the dynamic HF VS variables dominated with regard to importance in model performance. In classifying suspected (and tested) versus not suspected (and not-tested) patients, HR and RR features (variability, mean values, and monitoring frequency) were identified by the S+/- models as being important predictive features at both time prediction points (T_0 and T_{-24}). Interestingly, in both models using high frequency VS data, multiple HR and RR mean values of differing durations (e.g., 3, 12 and 24 h) feature characteristics were independently predictive, suggesting that both HR and RR alone and their trends over time are relevant characteristics of VTE phenotype. In models predicting HA-VTE diagnosis, BP features were identified as important to HA-VTE risk prediction 24 hours in advance of diagnostic test, but closer to the time of the actual outcome event, HR and RR features became

more important predictors. These findings have important implications for clinicians. Tachycardia and tachypnea are known to be associated with HA-VTE risk, yet they can also be common and non-specific findings in acutely ill, hospitalized patients. Furthermore, discerning subtle trends in vital sign patterns over a period of 24 to 48 hours can be challenging when there are competing demands for clinician attention as well as multiple hand-offs in patient care due to shift changes. Our findings seem to suggest that placing dynamic VS features in models which include both static and LF data may help clinicians to identify at risk patients better than can be accomplished with their own clinical evaluation of trends. These findings warrant further exploration and underscore recommendations from prior studies that recommend exploration of more personalized risk factors and of cumulative and ongoing re-evaluations of these factors over time.[119, 126, 130]

3.6.3 Cost-Benefit Trade-off Considerations

As illustrated by the performance capabilities of current, commonly used VTE RAM's (section 3.6.1), all risk tools demonstrate sensitivity and specificity trade-offs. An awareness of the indications and performance strengths and weakness of a clinical tool is required by clinicians in order to use them appropriately. We previously discussed our best performing models when we focused on a single specific metric of either sensitivity or specificity, and the clinical benefits to each. To further explore cost benefit realities within individual models, we examined TPR/FPR and TNR/FNR trade-offs.

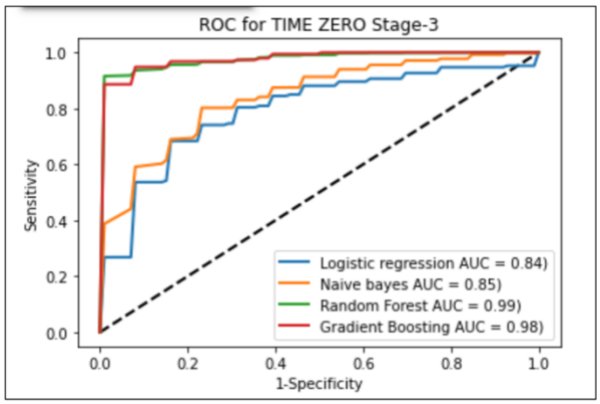
True positive means we correctly identify patients that are confirmed to have a positive HA-VTE diagnosis. False positive rate is the cost we bear when we identify a patient as having HA-VTE diagnosis, but they are actually negative for the disease. For this FTR diagnosis, which

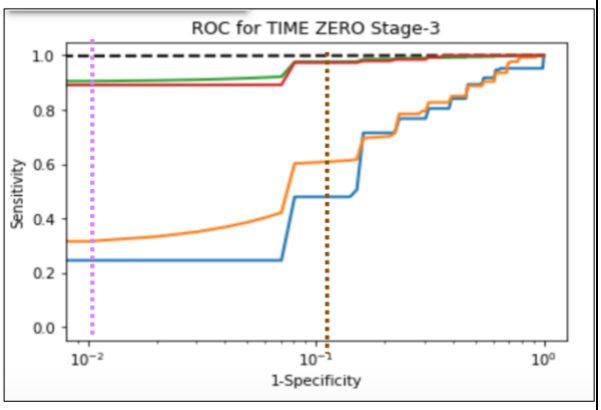
can be lethal if missed, we are more willing to accept an increase in false positives in order to maximize confidence of model ability to identify all cases that are truly positive.

To further explore cost and benefit, Figure 18 shows the area under the receiver operating characteristic (ROC) curve for all 4 classifiers when applied to C+/- models at T_0 (Panel A) and T_{-24} (Panel C). Panels B and D allow us to zoom into the left side of ROC curve and replot the ROC diagram showing the horizontal axis in base 10 logarithm scale. This allows us to examine the TPR at very low FPR to further examine the trade-offs.

When we look at TPR/FPR performance tradeoff, RF (green) has the best performance. At a false positive rate of 1% (dashed purple line in Panel B), the C+/- RF model can identify >80% of patients who will test positive for HA-VTE at T_0. If we increase our tolerance of false positives to 10% (dashed brown line), then that rate of true positive identification increases to 95% for RF (green) and GBDT (red) models and to 60% for the NB model (orange). As these are patients already suspected by clinicians of VTE risk and more likely to be receiving more intensive surveillance and pharmacologic prophylaxis, our cost of missing a positive case is high and we are more willing to accept a higher false positive rate in order to prevent a lethal complication. Similar cost-benefit tradeoffs are seen in the models predicting outcome at $T_{-24,}$ 24 hours in advance of event.

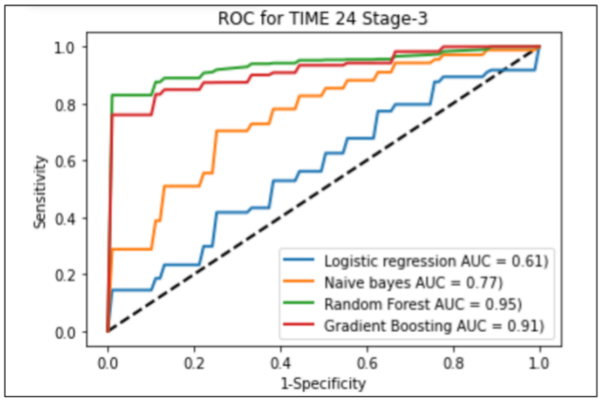
None of our individual models were equally good at both sensitivity and specificity at the various prediction timepoints. This work is early, but these cost-benefit trade-off evaluations can help to inform advancing this work toward the ultimate goal of developing a suite of models capable of providing the best sensitivity and the best specificity at the different timepoints for the different cohorts of interest.

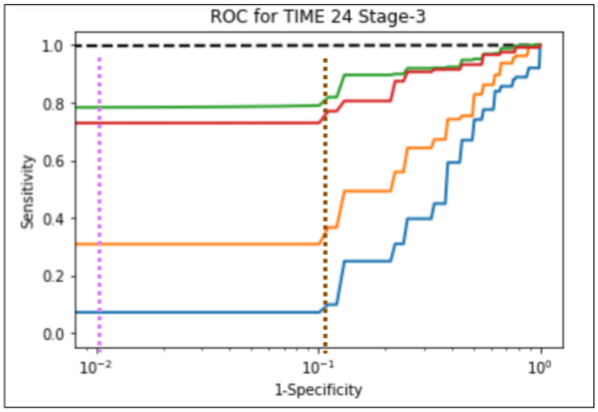




Panel A	Panel B

AUROC for Stage 3 C+/- Models using static data and dynamic lab, medication, and vital sign data accrued during a 24-hour observation window directly preceding T_0 (time of diagnostic test)





Panel C	Panel D

AUROC for Stage 3 C+/- Models using static data and dynamic lab, medication, and vital sign data accrued during a 24-hour observation window directly preceding T_{-24} (24-hrs before diagnostic test is conducted)

Figure 18. Area Under the Receiver Operating Characteristic (AUROC) Curve Diagrams displaying 10-fold cross validation results for discrimination between HA-VTE positive and HA-VTE negative (disease free) diagnosis. Sensitivity (TPR) is on the y-axis. FPR (1-Specificity) is on the horizontal axis, plotted at normal linear scale in Panels A and C, and base 10 logarithmic scale in Panels B and D

3.6.4 Comparison of Study Findings to Current Literature Exploring VTE Risk Prediction Through ML Approaches

ML approaches using classification algorithms have been used to successfully develop highly sensitive models to predict risk of cardiorespiratory instability, cardiac ischemia and arrhythmia,[81-83] and inform decision making to improve care delivery and patient outcomes. Studies exploring VTE risk assessment using ML methods are reported in the literature, but are however limited, and interpretation of results must be done through a close examination of study methodologies. The ability to develop models that can be best translated to clinical practice necessitates constructing models based on data sets that reflect real-world readily available clinical data elements and population incidence of VTE disease, which vary from 1-16% in medical patients to as high as 40% in certain surgical and oncologic populations. [36, 46]

In chemotherapy treated ambulatory cancer patients, Ferroni et. al, compared machine learning approaches with the Khorana model, a VTE RAM developed for outpatient populations with diagnosed malignancy. Ferroni's models highlighted variables not included in the Khorana model, however overall ML model performance was poor, and they acknowledge this was likely due to the study design.[131] Sample size was small (n=40) and the training platform used for model development was constructed by 1:1 matching of VTE cases with controls, however, VTE prevalence within their testing set was only 8%. James et al , used RF algorithms to leverage EHR data to predict the occurrence of HA-VTE.[132] Models were developed from a training data set of 1089 cancer patients (a population at high risk for VTE pathology) with VTE who were matched to 1089 control patients. Their results lacked validation by a clinical data set that conformed to actual incidence of VTE in hospitalized patients. Both of these studies based their evaluation of model performance on AUROC, a metric that works well when data is balanced, but one that does

not provide the most optimal insight into predictive performance for models developed to identify disease pathology that can be fatal when missed.

Kawaler et el, used ML methods (RF, NB, support vector machines, k-nearest neighbor, and C4.5 decision tree induction) to develop models predictive of post-hospitalization VTE using information automatically elicited from the EHR.[133] Strengths of their approach included a large sample size obtained from a clinical health care system. Their data set was unique in that over 60% of the sample patients had 20-plus years of clinical data available within the hospital's clinical data archive system and the majority of the patients in their sample had genotype data included in their EHR data as part of the eMERGE network, thus they were able to include 32 single nucleotide polymorphisms (SNP) associated with VTE in addition to demographic, diagnoses, labs, medications, procedures and vital signs. Similar to the approach used in this study, they evaluated their model performance using precision-recall metrics and the NB and RF classifiers showed the strongest performance in predicting HA-VTE risk post-hospital discharge. Their models' ability to predict post-hospital VTE risk outperformed existing risk assessment models and underscore the impact that leveraging routinely collected EHR data can have for informing and improving VTE risk prediction. Their model building methods were rigorous, and their results impressive and worthy of further exploration. However, it should be noted that most health care systems do not currently have the resources to support the infrastructure required for maintain e data warehouses that make decades of clinical and comprehensive genotype data readily available for routine risk prediction. As well, we believe it is worthy to identify models which can predict risk of HA-VTE while patients are still being cared for the in the hospital setting with enough lead time to implement targeted prophylaxis, or supportive measures in advance of an acute in-hospital

mortality event. Likewise, being able to identify those not at risk could spare low-risk patients from exposure to unnecessary prophylaxis and side effects

3.7 Limitations

3.7.1 Sample Limitations

While our sample size allows for diversity of diagnoses, variables, and adequate VTE prevalence to support machine learning, we recognize that generalizability of findings will be limited due to the fact that the sample population is restricted to a single SDU over a specific time interval, and that sample ethnicity is primarily White.

The high frequency VS data used in model development was only available during the time that patients in this sample were in the SDU, and accessed for the parent study, and thus our ability to explore HA-VTE risk using these dynamic data for modeling, was limited to this time frame. The prevalence of some clinical feature variables in our sample population during their SDU stay was low. This is due in part to the manner in which these data were identified and accessed for the parent study. As a result, many desired variable categories needed to be excluded due to absence of information that would only add noise and bias to model development. This reality has provided valuable insights that will guide data collection strategies and rationales in future work. A benefit of this limitation is that we have developed models that, using dynamically accruing VS data and readily available clinical data, appear to out-perform models in the literature which rely primarily on static data. With future development and testing, our modeling approach shows preliminary

promise, and can be further explored with more EHR feature heavy models to determine the optimal trade-offs between performance and parsimony, and potential clinical utility.

3.7.2 Ground Truth Limitations

We selected *time the gold standard radiology test was conducted* as the proxy time for ground truth identification of HA-VTE event outcome in this study. We are aware the ground truth does not necessarily equal time of diagnostic testing, or even time diagnosis or of clinician suspicion of diagnosis. Because of the presence of underlying VTE pathology, some positive cases may have been positive for hours to days in advance of our annotated ground truth time. Furthermore, the "not tested cases" may contain HA-VTE unrecognized by clinicians, and that these unrecognized positive cases may not be diagnosed until after this hospital admission. However, we were able to build a set of models with confirmed negative cases as our controls and ideas for future work include developing models with annotated outcome event defined as the time a gold standard radiologic test for HA-VTE is ordered by clinicians, to better approximate time disease is suspected to be present.

3.7.3 Model Development Limitations

This dissertation study took a survey approach to model development. A variety of classification algorithms were employed. We explored model development with a variety of sample cohorts, prediction time windows, and temporal clinical data, and we focused on training and testing models using cross-validation methods. Future planned work to build upon these initial models includes:

- Determining the optimal set points of individual models for cost-benefit considerations
- Hyperparameter tuning to enhance model performance at chosen set points
- Evaluating model performance with an external data set

3.7.4 Prediction Time Windows Limitations

As this is the first study to develop prediction models for new onset HA-VTE using data accrued during hospitalization, we recognize that the time windows proposed may not be the most optimal time frames for predicting this dynamic disease pathology. Predicting HA-VTE at time of diagnosis, while not meaningful for changing current clinical practice, is a requisite first step in temporal risk modeling research that aims to identify patient features associated with the evolution of HA-VTE complication development and risk. There are no published time interval recommendations for diagnostic VTE surveillance screening in the SDU population. The chosen prediction times represent an initial starting point that informed the aims of this dissertation study while also establishing the groundwork for future lines of inquiry that will explore additional time windows in advance of diagnosis.

3.7.5 Study Timeframe Limitations

Another limitation is that this study relied on data from a parent study, with data accrual occurring between 11/06 and 09/08. Little has changed in diagnostic testing for VTE, or VS data accrued from bedside monitors, however, conducting the study in a more recent timeframe would

allow for more current diagnosis and procedure coding methods, as well as more contemporary EHR data identification, storage and extraction methods. Every effort was made to map clinical data from this sample to current data standards (such as mapping ICD-9 codes to ICD-10 codes) and to include those mappings in manuscripts associated with this work, to maximize contemporary understanding and application.

3.8 Strengths

This study has several strengths. Many prior VTE risk studies are limited in that they report lack of confidence that their control cohorts are free of unidentified VTE cases. A strength of our study is that we used a rigorous and reproducible method for identifying confirmed HA-VTE positive, confirmed negative HA-VTE cases, and those confirmed not tested during hospital length of stay. The low prevalence of HA-VTE used in our model development is another strength. We trained models on data that reflected real-world population rates of HA-VTE in a total hospital unit census, with rates of 32% in clinical rationale #1 models and 6.42% in the second set of clinical models. Our approach developed models that aimed to predict HA-VTE risk in time windows in advance of outcome that could one day support clinician ability to intervene and potentially change outcomes. Finally, these are the only HA-VTE models which use data accruing dynamically during hospitalization, and thus follow the HA-VTE evolution timeline.

3.9 Conclusion

This study demonstrates that applying classification algorithms to leverage routinely collected, intensively collected time series clinical data has promise to predict new-onset HA-VTE risk in hospitalized patients at time of, and in advance of, diagnosis. The incorporation of high frequency dynamic patient vital sign data improves model performance compared with static data models. Furthermore, our results demonstrated that clinical features predictive of HA-VTE change in importance over time course of hospitalization. Our initial findings set the stage for further work and demonstrate that models which can better predict which hospitalized patients require more intense HA-VTE surveillance and prophylaxis, and also identify low-risk patients who can be spared unneeded prophylaxis are possible in future. The ultimate goal is to improve HA-VTE risk prediction approaches and better inform nursing surveillance and clinical decision making and at the bedside.

Appendix A. Dissertation Manuscript 1

ICU Scoring Systems

Abstract

Background. Severity scoring systems are commonly used in critical care and, when applied to the populations for whom they were developed and validated, these tools can inform mortality prediction and risk stratification, resource utilization, and optimization of patient outcomes.

Methods. Original articles published in the English language were identified through OVID and MEDLINE literature searches conducted for the years 1980 to 2020. A list of terms associated with critical care scoring systems were used alone or in combination for the literature search.

Results. This article appraises the characteristics and applications of the scoring systems most frequently applied to critically ill patients: those that predict risk of in-hospital mortality at time of ICU admission (APACHE, SAPS, and MPM), and those that assess and characterize current degree of organ dysfunction (MODS, SOFA, and LODS). Variable type and collection timing, score calculation, patient population, and comparative performance data of these systems are detailed.

Conclusion: Awareness of the strengths, limitations, and specific characteristics of severity scoring systems commonly used in ICU patients is vital for critical care nurses to effectively employ these tools in clinical practice and to critically appraise research findings based on their usage.

Permission to reproduce the abstract only of Pellathy T, Pinsky MR, Hravnak M. ICU Scoring Systems. *Critical Care Nurse* (in press), was granted by Michael Muscat, AACN Publishing Manager and is included below. The full text article can be accessed via the journal's website, https://aacnjournals.org/ccnonline.

AMERICAN ASSOCIATION of CRITICAL-CARE NURSES

November 7, 2020

Tiffany Purcell Pellathy, MS, ACNP-BC
University of Pittsburgh School of Nursing
3500 Victoria Street
Pittsburgh, PA 15213

Dear Ms. Pellathy:

Thank you for your reuse request. We hereby grant permission for your reuse of the AACN copyrighted content below, free of charge, subject to the following conditions:

1. Content will be made available on an electronic thesis database at the University of Pittsburgh School of Nursing.
2. Permission is granted for the abstract only of the in-press manuscript CCN-D-19-00073R2. Once the manuscript is published, a hyperlink to the online version will be placed in the electronic thesis database with the abstract. The online version will require a subscription for full access.
3. Full and suitable acknowledgment to the original source must be made in a prominent place along with the content included in the thesis database.
4. Permission is granted for the following use case: full manuscript, educational use, website (login required), United States, original language, no more than 499 viewers, no modifications, current edition and up to 5 years (until November 7, 2025).

Thank you for your interest in the American Association of Critical-Care Nurses.

Sincerely,

Michael Muscat
AACN Publishing Manager

Accepted:

SIGNATURE	TITLE	DATE
	Pre-doctoral fellow	11/07/2020

Appendix B. Dissertation Manuscript 2

This manuscript, **Accuracy of Identifying Venous Thromboembolism by Administrative Coding: Implications for Big Data and Machine Learning Research,** has been accepted at the time of ETD submission, for publication in the *Journal of Clinical Monitoring and Computing*.

Authors: Tiffany Pellathy MS, ACNP-BC[1], Melissa Saul, MS[2], Gilles Clermont MD, MSC[2], Artur W. Dubrawski PhD[3], Michael R. Pinsky MD[2], Marilyn Hravnak RN, PhD[1]

Affiliations: University of Pittsburgh School of Nursing[1,] University of Pittsburgh School of Medicine[2], and Carnegie Mellon University, School of Computer Science, Auton Lab[3]

ABSTRACT

Purpose: Big data analytics research using heterogeneous electronic health record (EHR) clinical data requires accurate identification of disease phenotype cases and controls. Hospital-acquired venous thromboembolism or pulmonary embolism (HA-VTE) is particularly challenging to identify due to its temporal evolution over hospitalization and variable EHR documentation. To establish ground truth for machine learning modeling, we compared the accuracy of HA-VTE diagnoses made by administrative coding to manual review of gold standard diagnostic test results.

Methods: We performed retrospective analysis of EHR data on 3680 adult stepdown unit patients identifying HA-VTE. International Classification of Diseases, Ninth Revision (ICD-9-CM) codes for VTE were identified. 4455 radiology reports associated with VTE diagnostic tests were

screened using terminology extraction and then manually reviewed by a clinical expert to confirm diagnosis. Clinical notes were reviewed to clarify indeterminate results.

Results: Of 415 cases with ICD-9-CM codes for VTE, 219 were identified with acute onset type codes. Test report review identified 132 new-onset HA-VTE cases. Only 40% of ICD-9-CM coded cases (n=87) were confirmed by a positive diagnostic test report, leaving the majority of administratively coded cases unsubstantiated by confirmatory diagnostic test. Additionally, 54% of HA-VTE cases confirmed by diagnostic test lacked corresponding codes.

Conclusions: ICD-9-CM coding missed diagnostic test-confirmed HA-VTE cases and inaccurately assigned cases without confirmed VTE, suggesting dependence on administrative coding leads to inaccurate HA-VTE phenotyping. Alternative methods to develop more sensitive and specific VTE phenotype solutions portable across EHR vendor data are needed to support case-finding in big-data analytics applied to this population.

Keywords: Administrative coding, Venous Thromboembolism, big data analytics, electronic health record data, phenotyping, machine learning

Appendix C. Drug Classification Table

A breakdown of the classification of the drug variables listed in Table 4 can be found in the table below.

Table 12. Medication Categories

Medication	Drug category
Antithrombin_III	anti-coagulant
Argatroban	anti-coagulant
Bivalirudin	anti-coagulant
Argatroban	anti-coagulant
Dalteparin	anti-coagulant
Enoxaparin	anti-coagulant
Fondaparinux	anti-coagulant
Fragmin	anti-coagulant
Heparin	anti-coagulant
Warfarin	anti-coagulant
Amikacin	anti-infective
Amoxicillin	anti-infective
Ampicillin	anti-infective
Dicloxacillin	anti-infective
Nafcillin	anti-infective
Oxacillin	anti-infective
Penicillin	anti-infective
Piperacillin	anti-infective
Piperacillin/tazobactam	anti-infective
Ticarcillin/clavulanate	anti-infective
Amphotericin B	anti-infective
Atazanavir	anti-infective
Augmentin	anti-infective
Azithromycin	anti-infective
Aztreonam	anti-infective
Caspofungin	anti-infective
Cefazolin	anti-infective
Cefdinir	anti-infective
Cefuroxime	anti-infective

Ceftazidime	anti-infective
Cephalexin	anti-infective
Clindamycin	anti-infective
Cefazolin	anti-infective
Cefdinir	anti-infective
Cefepime	anti-infective
Cefotaxime	anti-infective
Cefotetan	anti-infective
Cefpodoxine	anti-infective
Ceftazidime	anti-infective
Ceftriaxone	anti-infective
Cefuroxime	anti-infective
Doxycycline	anti-infective
Efavirenz	anti-infective
Famciclovir	anti-infective
Itraconazole	anti-infective
Ketoconazole	anti-infective
Piperacillin/Tazobactam	anti-infective
Sulfamethoxazole/Trimethoprim	anti-infective
Valganciclovir	anti-infective
Unasyn (ampicillin/sulbactam)	anti-infective
Gentamicin	anti-infective
Vancomycin	anti-infective
Moxifloxacin	anti-infective
Abacavir	anti-infective
Amoxicillin	anti-infective
Ampicillin	anti-infective
Dicloxacillin	anti-infective
Nafcillin	anti-infective
Oxacillin	anti-infective
Penicillin	anti-infective
Piperacillin	anti-infective
Ticarcillin	anti-infective
Aspirin	anti-platelet
Clopidogrel	anti-platelet
Eptifibatide	anti-platelet
Chlorpromazine	anti-psychotic
Dacarbazine	anti-psychotic
Fluphenazine	anti-psychotic

Prochlorperazine	anti-psychotic
Promethazine	anti-psychotic
Ranolazine	anti-psychotic
Thioridazine	anti-psychotic
Droperidol	anti-psychotic
Aripiprazole	anti-psychotic
Quetiapine	anti-psychotic
Olanzapine	anti-psychotic
Haloperidol	anti-psychotic
Risperidone	anti-psychotic
Perphenazine	anti-psychotic
Chlordiazepoxide	anti-psychotic
Ziprasidone	anti-psychotic
Divalproex	anti-psychotic
Lithium	anti-psychotic
Nortriptyline	anti-psychotic
Amitriptyline	anti-psychotic
Lorazepam	benzodiazepine
Diazepam	benzodiazepine
Alprazolam	benzodiazepine
Adenosine	cardiac
Amiodarone	cardiac
Amlodipine	cardiac
Atenolol	cardiac
Atorvastatin	cardiac
Atropine	cardiac
Amlodipine	cardiac
Carvedilol	cardiac
Betaxolol	cardiac
Bisoprolol	cardiac
Clonidine	cardiac
Digoxin	cardiac
Diltiazem	cardiac
Dobutamine	cardiac
Enalapril	cardiac
Ezetimibe/simvastatin	cardiac
Diovan	cardiac
Dopamine	cardiac
Esmolol	cardiac

Fluvastatin	cardiac
Fosinopril	cardiac
Hydralazine	cardiac
Hydrochlorothiazide	cardiac
Irbesartan	cardiac
Isosorbide	cardiac
Labetalol	cardiac
Lisinopril	cardiac
Milrinone	cardiac
Nadolol	cardiac
Nicardipine	cardiac
Nitroglycerin	cardiac
Nitroprusside	cardiac
Olmesartan	cardiac
Pravachol	cardiac
Pravastatin	cardiac
Quinapril	cardiac
Ramipril	cardiac
Rosuvastatin	cardiac
Sildenafil	cardiac
Spironolactone	cardiac
Verapamil	cardiac
Tenoretic (atenolol/chlorthalidone)	cardiac
Vinblastine	chemo
Vincristine	chemo
Darbepoetin	erythro-stimulant
Erythropoietin	erythro-stimulant
Famotidine	GI
Esomeprazole	GI
Lansoprazole	GI
Octreotide	GI
Omeprazole	GI
Ranitidine	GI
Sulfasalazine	GI
Vonwillebrand/ahf-factor	hemostasis
Anastrozole	hormone
Estradiol	hormone
Levothryoxine	hormone
Premarin (estrogens, conjugated)	hormone

Prempro (estrogens, conjugated/medroxyprogesterone)	hormone
Tamoxifen	hormone
Testosterone	hormone
Glimepiride (sulfonylurea)	hyperglycemic
Glipizide (sulfonylurea)	hyperglycemic
Glucagon	hyperglycemic
Glyburide (sulfonylurea)	hyperglycemic
Humalog	hyperglycemic
Insulin	hyperglycemic
Repaglinide (meglitinide)	hyperglycemic
Rosiglitazone (thiazolidinediones, TZDs)	hyperglycemic
Sitagliptin (DPP-4 inhibitor)	hyperglycemic
Tacrolimus	immuno-suppressant
Rituximab	immuno-suppressant
Rasburicase	immuno-suppressant
Cyclosporine	immuno-suppressant
Rituximab	immuno-suppressant
Interferon	immuno-suppressant
Naloxone	naloxone
Meperidine	narcotic
Acetaminophen/codeine	narcotic
Morphine	narcotic
Oxycodone	narcotic
Oxycodone	narcotic
Percocet	narcotic
Vicodin	narcotic
Fentanyl	narcotic
Hydromorphone	narcotic
Codeine	narcotic
Amantadine	neurologic
Carbamazepine	neurologic
Carbidopa/levodopa	neurologic
Gabapentin	neurologic
Phenobarbital	neurologic
Phenytoin	neurologic
Sinemet	neurologic
Gabapentin	neurologic
Levetiracetam	neurologic
Etodolac	NSAID

Ibuprofen	NSAID
Ketorolac	NSAID
Celecoxib	NSAID
Acetaminophen	pain
Zolpidem	sedative
Clonazepam	sedative
Propofol	sedative
Sertraline	SSRI
Venlafaxine	SSRI
Citalopram	SSRI
Duloxetine	SSRI
Fluoxetine	SSRI
Paroxetine	SSRI
Escitalopram	SSRI
Dexamethasone	steroids
Hydrocortisone	steroids
Methylprednisolone	steroids
Methylpred	steroids
Prednisolone	steroids
Prednisone	steroids
Alteplase	thrombolytic
Reteplase	Thrombolytic
Dopamine	vasopressor
Epinephrine	vasopressor
Norepinephrine	vasopressor
Vasopressin	vasopressor

GI *gastrointestinal*, NSAID *non-steroidal anti-inflammatory drugs*, SSRI *selective serotonin reuptaker inhibitor*

Appendix D: Mapping of Study Variables to Existing Risk Assessment Models

Table 13. Study Variables Mapped to Existing VTE Risk Assessment Models

Risk Assessment Model	IMPROVE (DD)	GENEVA	PADUA	CAPRINI	KUCHER	Cook D et. al.	Inclusion of established RAM variables in this research study (S: static variable; LF: low frequency variable):		
Validated hospital population	Medical	Medical	Medical/ Surgical	Medical/ Surgical	Medical/ Surgical	ICU Surgical			
Sensitivity	63.3	75	61.8	88.6	28	n/a			
Specificity	70.7	34.1	48.8	21.4	85.7	n/a	**Variable name**	**Data source**	**Details**
Any prior VTE	X	X	X	X	X		Prior VTE	ICD-9, radiology	Detailed in section 1.3.2.1
Thrombophilia	X		X				Known clotting disorder	ICD-9, lab results	S: 286.x; 289.81 (primary); 289.82 (secondary) LF: Individual lab test results
Known hypercoagulable state		X			X				
Factor V Leiden*				X					
Prothrombin 20210A*				X					
(↑) Serum homocysteine*				X					
(+) lupus anticoagulant*				X	X				
(↑) Anti-cardiolipin antibody*				X	X				
HITT				X					
Elevated D-dimer twice the upper limits of normal	X								
Active malignancy	X	X	X		X	X	Malignancy (past or present)	ICD-9	S: 140.x-208.x and 230.x-234.x
Malignancy past/present					X				

Risk Assessment Model	IMPROVE (DD)	GENEVA	PADUA	CAPRINI	KUCHER	Cook D et. al.	Inclusion of established RAM variables in this research study (S: static variable; LF: low frequency variable):		
Acute infection/sepsis	X	X		X			Infection/sepsis	ICD-9, lab results	S: 00x.x-139.x, 785.5, 790.7
Age > 60 years	X	X		X			Age	Date of birth (DOB)	S: Calculated from DOB and used as a continuous variable
Age > 70 years			X		X				
BMI > 25				X			Obesity	ICD-9	S: 278.0
BMI>29 or obesity dx					X				
Obesity						X			
BMI ≥ 30			X						
Central venous catheter (CVC)				X			CVC	Charge data	LF
Lung disease				X			Chronic Pulmonary Disease	ICD-9	S: 490.x-492.x, 493.9, 496, 460.x-519.x
COPD						X			
Stroke within past 90 days		X					Prior Stroke	ICD-9	S: 438.x
Stroke				X		X			
Acute ischemic stroke			X				Acute stroke	ICD-9	S: 434.x
Myocardial infarction (MI) within 90 days		X					Prior MI	ICD-9	S: 412.x, 414.x
Acute MI			X	X		X	Acute MI	ICD-9	S: 410.x; 411.x
Congestive heart failure (CHF)		X	X	X		X	CHF	ICD-9	S: 428.x
Diabetes mellitus (DM)				X		X	DM, no complications	ICD-9	S: 250-250.3, 250.8x, 250.9x
							DM, with complications	ICD-9	S: 250.4-7
Chemotherapy*							Chemo	Medication	LF: anti-neoplastic

Risk Assessment Model	IMPROVE (DD)	GENEVA	PADUA	CAPRINI	KUCHER	Cook D et. al.	Inclusion of established RAM variables in this research study (S: static variable; LF: low frequency variable):		
Fracture hip/leg						X	Lower extremity (LE) fracture	ICD-9	S: 820.x-823.x
Plaster cast				X					
Trauma (multiple)				X			Trauma	ICD-9	S: 800.x-959.x
Trauma (any)			X						
Acute SCI with paralysis				X			SCI	ICD-9	S: 952.x
Varicose veins				X			Varicose Veins	ICD-9	S: 454.x
Chronic venous insufficiency		X					Chronic venous insufficiency	ICD-9	S: 459.x
Irritable Bowel Syndrome (IBS) history				X			IBS	ICD-9	S: 564.1
Hormone therapy		X	X		X		Hormone therapy	Medication	LF: hormone meds
Hypertension (HTN)						X	HTN	ICD-9, meds	S: 401.x, 402.x LF: Cardiac meds
Major surgery (> 60 min)					X		Major surgery (> 45 min)	DRG	LF:
Major surgery (> 45 min)				X			Minor surgery (< 45 min)		
Minor surgery (< 45 min)				X			LE elective arthroplasty		
LE elective arthroplasty				X					
Surgery type (GI, CABG, vascular, orthopedic, trauma, neurosurgery, thoracic, GYN malignancy)							Surgery type (GI, CABG, vascular, orthopedic, trauma, neurosurgery,	DRG	LF

Risk Assessment Model	IMPROVE (DD)	GENEVA	PADUA	CAPRINI	KUCHER	Cook D et. al.		Inclusion of established RAM variables in this research study (S: static variable; LF: low frequency variable):		
								thoracic, GYN malignancy)		
Abnormal lab values (Plt, Hgb, WBC, HgbA1c, Glucose, HDL, LDL)								Lab values	Lab data	LF: See Table 4, Section 1.4.4.2 for all lab values included
Variables that could not be reliably represented in the study data										
Recent surgery								Unable to identify in this data set.		
Respiratory failure								Unable to reliably capture in this SDU patient population.		
Reduced mobility								LE fracture and trauma are captured in other variables.		
LE paralysis during hospitalization								Stroke/hemiparesis and LE immobilization are captured in other variables.		
Immobilization ≥ 7 days								Unable to identify in this data set.		
Pregnancy*								Pregnant patients were not excluded from this sample, but no diagnosis of pregnancy was found within this sample.		
Dehydration*								Unable to identify in this data set.		
Swollen legs								Unable to identify in this data set.		
Family history of thrombosis								Unable to identify in this data set.		
Recent travel > 6 hours								Unable to identify in this data set.		
Nephrotic syndrome*								No diagnoses for nephrotic syndrome in this sample population and no other means of identifying it reliably.		
Myeloproliferative syndrome*								Diagnosis codes checked and no codes associated with this diagnosis were found.		
Acute Rheumatic disease*								Diagnosis codes checked and no codes associated with this diagnosis were found.		

* Diagnosis/test/status not found in this sample during data screening and cleaning

Appendix E: Institutional Review Board (IRB) Approval

From: <irb@pitt.edu>
Date: June 14, 2018 at 1:10:02 PM EDT
To: <tdp20@pitt.edu>
Subject: PI Notification: Your research study received approval under expedited review
Reply-To: <irb@pitt.edu>

University of Pittsburgh
Institutional Review Board

3500 Fifth Avenue
Pittsburgh, PA 15213
(412) 383-1480
(412) 383-1508 (fax)
http://www.irb.pitt.edu

Memorandum

To:	Tiffany Pellathy ,
From:	Sue Beers, Vice Chair
Date:	6/14/2018
IRB#:	PRO18050409
Subject:	Machine Learning to Determine Dynamically Evolving New-Onset Venous Thromboembolic (VTE) Event Risk in Hospitalized Patients

The University of Pittsburgh Institutional Review Board reviewed and approved the above referenced study by the expedited review procedure authorized under 45 CFR 46.110 and 21 CFR 56.110. Your research study was approved under:

45 CFR 46.110.(5)

The risk level designation is Minimal Risk.

Approval Date: 6/14/2018
Expiration Date: 6/13/2019

For studies being conducted in UPMC facilities, no clinical activities can be undertaken by investigators until they have received approval from the UPMC Fiscal Review Office.

Please note that it is the investigator's responsibility to report to the IRB any unanticipated problems involving risks to subjects or others [see 45 CFR 46.103(b)(5) and 21 CFR 56.108(b)]. Refer to the IRB Policy and Procedure Manual regarding the reporting requirements for unanticipated problems which include, but are not limited to, adverse events. If you have any questions about this process, please contact the Adverse Events Coordinator at 412-383-1480.

The protocol and consent forms, along with a brief progress report must be resubmitted at least one month prior to the renewal date noted above as required by FWA00006790 (University of Pittsburgh), FWA00006735 (University of Pittsburgh Medical Center), FWA00000600 (Children's Hospital of Pittsburgh), FWA00003567 (Magee-Womens Health Corporation), FWA00003338 (University of Pittsburgh Medical Center Cancer Institute).

Please be advised that your research study may be audited periodically by the University of Pittsburgh Research Conduct and Compliance Office.

University of Pittsburgh
Institutional Review Board

Human Research Protection Office
3500 Fifth Avenue, Suite 106
Pittsburgh, PA 15213
Tel (412) 383-1480
www.hrpo@pitt.edu

APPROVAL OF CONTINUING REVIEW/TRANSITION (Expedited)

Date:	June 17, 2019
IRB:	STUDY19050329
PI:	Tiffany Pellathy
Title:	Machine Learning to Determine Dynamically Evolving New-Onset Venous Thromboembolic (VTE) Event Risk in Hospitalized Patients
Funding:	Name: National Institute of Nursing Research, Grant Office ID: 05.32207.xxxx.00000.131227.00000, Funding Source ID: 1F31NR018102
Grant ID:	05.32207.xxxx.00000.131227.00000;

The Institutional Review Board reviewed and approved the above referenced study. The study may continue as outlined in the University of Pittsburgh approved application and documents.

Approval Documentation

Review type:	Initial Study
Approval Date:	6/17/2019
Approved Documents:	• RESEARCH_STRATEGY, Category: IRB Protocol; • NOTICE OF AWARD, Category: Sponsor Attachment; • HUMAN_Pellathy.docx, Category: Waiver Script;

As the Principal Investigator, you are responsible for the conduct of the research and to ensure accurate documentation, protocol compliance, reporting of possibly study-related adverse events and unanticipated problems involving risk to participants or others. The HRPO Reportable Events policy, Chapter 17, is available at http://www.hrpo.pitt.edu/.

Clinical research being conducted in an UPMC facility cannot begin until fiscal approval is received from the UPMC Office of Sponsored Programs and Research Support (OSPARS).

If you have any questions, please contact the University of Pittsburgh IRB Coordinator, Larry Ivanco.

Please take a moment to complete our Satisfaction Survey as we appreciate your feedback.

Appendix F. Responsible Conduct of Research Activities

Table 14. Responsible Conduct of Research Training Activities

Date	Activity Name	Presenter	Subject Matter	Format & Duration
January-August 2020	**Mentored Research Apprenticeship**	Marilyn Hravnak, PhD	mentor/mentee responsibilities and relationships; data acquisition and laboratory tools; management, sharing and ownership, responsible authorship and publication	1 hour, twice a week for 8 months
As part of this mentored apprenticeship, the PI participates and contribute to a weekly, 1-hour, multi-disciplinary research team meeting and also a weekly 1 hour meeting (one-on-one) with Dr. Hravnak. Multiple R01 funded projects are discussed, methodology outlined and critiqued, data analysis reviewed and refined and dissemination of findings is ongoing through collaborative writing of conference abstracts and manuscript preparation. Through this apprenticeship, the PI has advanced her knowledge and application of data science and machine learning approaches to understanding predictive patterns of disease. This past year she has focused on the logistics of waveform data extraction, working with computer programmers to create a user-friendly viewer for data annotation by a team of experts and on rigorous methods for ground truth annotation of clinical data for machine learning predictive analytics. She has continued to refine her ability to communicate domain specific information and to contribute domain expertise within a multidisciplinary group and she has collaborated on a successful national conference abstract submission and ongoing manuscripts.				
12/1/2019	**Data Science Ethics**	H.V. Jagadish, University of Michigan	Ethics	Online-course 12.5 hours
What are the ethical considerations regarding the privacy and control of consumer information and big data, especially in the aftermath of recent large-scale data breaches? This course provides a framework to analyze these concerns as you examine the ethical and privacy implications of collecting and managing big data. Explore the broader impact of the data science field on modern society and the principles of fairness, accountability and transparency as you gain a deeper understanding of the importance of a shared set of ethical values. You will examine the need for voluntary disclosure when leveraging metadata to inform basic algorithms and/or complex artificial intelligence systems while also learning best practices for responsible data management, understanding the significance of the Fair Information Practices Principles Act and the laws concerning the "right to be forgotten." This course will help you answer questions such as who owns data, how do we value privacy, how to receive informed consent and what it means to be fair.				
10/3/19	**An Author's Responsibilities: Publication and Authorship**	Robert Weyant	Responsible authorship	Face-to-face training session 1 hour

The author's role in writing, submitting, and ultimately publishing scientific research results ethically will be discussed. Perspectives of the first author, coauthors, journal editors, and other contributors to scientific publication will be addressed.				
09/20/19	**Data Sharing**	Helenmary Sheridan, MLIS Data Services Librarian	Data collection monitoring and reporting	Face-to-face training session 2 hours
Many funders, publishers, and institutions require researchers to make their research data public, but practical challenges can act as a barrier to sharing data, especially in the health sciences. This hands-on workshop will guide participants through the data sharing process, from initial study design to data deposit. Exercises will prompt participants to think through issues of data documentation, reuse value, and promotion of their own research projects				
7/25/19	**Introduction to Research Data Management**	Helenmary Sheridan, MLIS Data Services Librarian	Data acquisition	Face-to-face training session 2 hours
In this class, the fundamentals of keeping your data secure and organized are reviewed through brief introductions to the core areas of data management: file storage and organization, file documentation, data preservation, and data publication and/or data sharing. This class is intended for graduate students and researchers who are working on long-term research projects, or for anyone who wants to make sure their personal files are safe for the long-term				
5/20/19	**IRB Bootcamp**	Tonja M Hartjes, RN, PhD	Human Participants, Animal Subjects, Lab Safety	Face-to-face training session 2.5 hours
This presentation is designed to provide the APRN with information to determine when Institutional Review Board (IRB) is required and to assist with navigating through the IRB process. Various case studies are presented to the participants to determine whether IRB is required, the type of study and investigators' responsibilities while engaged in research.				
5/21/19	**Collecting Nursing Research Data 24 Hours a Day: Challenges, Lessons, and Recommendations**	Mary Lou Sole, PhD, RN, CCNS, Steven Talbert, PhD, RN, Melody Bennett, MN, RN, CCRN, Aurea Middleton, BSN, RN, Lara Deaton, BSN, RN, CCRN, and Daleen	data acquisition	Face-to-face training session; panel discussion 1 hour

		Penoyer, PhD, RN, CCRP		
Clinical studies may require testing of interventions around the clock (24/7). Best practices for implementing a study 24/7 are needed. This seminar will describe challenges in conducting a nursing intervention study around the clock (24/7), Identify strategies to facilitate conducting a study 24/7 and discuss the importance of planning, staffing, and continuous assessment of quality improvement opportunities when conducting a clinical trial.				
5/21/19	**Enrollment Challenges in Critical Care Nursing Research**	Mary Lou Sole, PhD, RN, CCNS, Steven Talbert, PhD, RN, Melody Bennett, MN, RN, CCRN, Aurea Middleton, BSN, RN, Lara Deaton, BSN, RN, CCRN, and Daleen Penoyer, PhD, RN, CCRP	data acquisition	Face-to-face training session; panel discussion 1 hour
Enrollment of participants is a key to a study's success. Specific strategies can be designed to address barriers and promote enrollment. Data from an ongoing clinical trial are used to describe success and challenges in enrolling patients in clinical research. Describe the importance of accurately determining enrollment projects for clinical studies; Discuss strategies that increase the likelihood of achieving successful enrollment of subjects; Describe enrollment data and challenges in an ongoing critical care clinical trial				
6/27/19	**Basics of Health Privacy**	Reid Cushman, PhD	Human Participants, Animal Subjects, Lab Safety	On-line CITI module 1 hour
Explain the basic privacy protections for health information provided by HIPAA and other legal-regulatory sources; Discuss the duties imposed on persons with access to health information in order to secure those privacy protections; Compare HIPAA's additional privacy protections for individually identifiable health data that are used for human subjects research, including authorizations, accountings of disclosures, etc.; Contrast situations where full HIPAA privacy protections are required, and those which can qualify for waivers, alterations, or exemptions with more limited requirement.				

6/27/19	**Health Privacy Issues for Researchers** Reid Cushman, PhD	Human Participants, Animal Subjects, Lab Safety	On-line CITI module 1 hour	This module discusses data protection requirements for human subjects research that creates, obtains, uses, or discloses health data, principally the protections that derive from the
10/1/17	**It's only a Model: What Can and Can't be Learned from Computational Models**	Tim Lezon, PhD	Collaborative Research	Face-to-face training session 1 hour
Like any research tool, computational modeling is most effective if the user has a good feel for its inner workings. Communication failure between quantitative and wet bench scientists can doom transdisciplinary projects. This session will address the strengths and limitations of computational models, and will explore strategies for effectively communicating across disciplines.				
11/1/17	High Stakes Consenting	Michael Green, MD, MPH, Karen Schmidt, PhD	Human Participants, Animal Subjects, Lab Safety	Face-to-face training session 1 hour
Vulnerable populations present a unique set of demands for clinical research. Using cases, we will discuss how to identify potentially vulnerable participants and develop strategies and techniques for designing a consent experience that supports a responsible and ethical research consent process.				

Bibliography

1. Silber, J.H., et al., *Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue.* Med Care, 1992. **30**(7): p. 615-29.
2. Silber, J.H., et al., *Failure-to-rescue: comparing definitions to measure quality of care.* Med Care, 2007. **45**(10): p. 918-25.
3. Needleman, J., et al., *Nurse-staffing levels and the quality of care in hospitals.* N Engl J Med, 2002. **346**(22): p. 1715-22.
4. Aiken, L.H., et al., *Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction.* Jama, 2002. **288**(16): p. 1987-93.
5. Clarke, S.P. and L.H. Aiken, *Failure to rescue.* Am J Nurs, 2003. **103**(1): p. 42-7.
6. Johnston, M.J., et al., *A systematic review to identify the factors that affect failure to rescue and escalation of care in surgery.* Surgery, 2015. **157**(4): p. 752-63.
7. Grosse, S.D., et al., *The economic burden of incident venous thromboembolism in the United States: A review of estimated attributable healthcare costs.* Thromb Res, 2016. **137**: p. 3-10.
8. Shahi, A., et al., *The Incidence and Economic Burden of In-Hospital Venous Thromboembolism in the United States.* J Arthroplasty, 2017. **32**(4): p. 1063-1066.
9. Cohoon, K.P., et al., *Costs of venous thromboembolism associated with hospitalization for medical illness.* Am J Manag Care, 2015. **21**(4): p. e255-63.
10. Ramos, J.D., et al., *The Khorana Score in Predicting Venous Thromboembolism for Patients With Metastatic Urothelial Carcinoma and Variant Histology Treated With Chemotherapy.* Clin Appl Thromb Hemost, 2016: p. 1076029616668405.
11. Kafeza, M., et al., *A systematic review of clinical prediction scores for deep vein thrombosis.* Phlebology, 2017. **32**(8): p. 516-531.
12. Manyika, J., *Big data: The next frontier for innovation, competition, and productivity.* http://www. mckinsey. com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation, 2011.
13. Chen, L., et al., *Using Supervised Machine Learning to Classify Real Alerts and Artifact in Online Multi-signal Vital Sign Monitoring Data.* Critical care medicine, 2016. **44**(7): p. e456-e463.
14. Needleman, J. and P.I. Buerhaus, *Failure-to-rescue: comparing definitions to measure quality of care.* Med Care, 2007. **45**(10): p. 913-5.
15. Needleman, J., E.T. Kurtzman, and K.W. Kizer, *Performance measurement of nursing care: state of the science and the current consensus.* Med Care Res Rev, 2007. **64**(2 Suppl): p. 10s-43s.
16. Rafferty, A.M., et al., *Outcomes of variation in hospital nurse staffing in English hospitals: cross-sectional analysis of survey data and discharge records.* International journal of nursing studies, 2007. **44**(2): p. 175-182.
17. Kelly, L. and D. Vincent, *The dimensions of nursing surveillance: a concept analysis.* J Adv Nurs, 2011. **67**(3): p. 652-61.

18. Aiken, L.H., et al., *Educational levels of hospital nurses and surgical patient mortality.* Jama, 2003. **290**(12): p. 1617-23.
19. Elmufdi, F. and C.R. Weinert, *Decreasing Failure-to-rescue Events in the Era of Rapid Response Systems.* Clinical Pulmonary Medicine, 2015. **22**(5): p. 223-229.
20. Aiken, L.H., S.P. Clarke, and D.M. Sloane, *Hospital staffing, organization, and quality of care: Cross-national findings.* Nurs Outlook, 2002. **50**(5): p. 187-94.
21. Mushta, J., L.R. K, and E. Andersen, *Failure to rescue as a nurse-sensitive indicator.* Nurs Forum, 2018. **53**(1): p. 84-92.
22. Jones, A. and M.J. Johnstone, *Inattentional blindness and failures to rescue the deteriorating patient in critical care, emergency and perioperative settings: Four case scenarios.* Aust Crit Care, 2016.
23. Verrillo, S.C. and B.D. Winters, *Review: Continuous Monitoring to Detect Failure to Rescue in Adult Postoperative Inpatients.* Biomed Instrum Technol, 2018. **52**(4): p. 281-287.
24. Hravnak, M., et al., *Causes of failure to rescue*, in *Textbook of rapid response systems*. 2017, Springer. p. 95-110.
25. Subbe, C., et al., *Validation of a modified Early Warning Score in medical admissions.* Qjm, 2001. **94**(10): p. 521-526.
26. Singer, M., et al., *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3).* Jama, 2016. **315**(8): p. 801-10.
27. Smith, G.B., et al., *The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death.* Resuscitation, 2013. **84**(4): p. 465-470.
28. McGaughey, J., et al., *Outreach and Early Warning Systems (EWS) for the prevention of Intensive Care admission and death of critically ill adult patients on general hospital wards.* Cochrane Database of Systematic Reviews, 2007(3).
29. Alam, N., et al., *The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review.* Resuscitation, 2014. **85**(5): p. 587-94.
30. Winters, B.D., et al., *Rapid response systems: a systematic review.* Critical care medicine, 2007. **35**(5): p. 1238-1243.
31. Ranji, S.R., et al., *Effects of rapid response systems on clinical outcomes: systematic review and meta-analysis.* Journal of hospital medicine: an official publication of the Society of Hospital Medicine, 2007. **2**(6): p. 422-432.
32. Schmid, A., et al., *Failure to rescue: a literature review.* J Nurs Adm, 2007. **37**(4): p. 188-98.
33. Meyer, G. and M. Lavin, *Vigilance: the essence of nursing.* Online Journal of Issues in Nursing, 2005. **10**(3).
34. Manojlovich, M. and A. Talsma, *Identifying nursing processes to reduce failure to rescue.* Journal of Nursing Administration, 2007. **37**(11): p. 504-509.
35. Kutney-Lee, A., E.T. Lake, and L.H. Aiken, *Development of the Hospital Nurse Surveillance Capacity Profile.* Res Nurs Health, 2009. **32**(2): p. 217-28.
36. Goldhaber, S.Z., *Risk factors for venous thromboembolism.* J Am Coll Cardiol, 2010. **56**(1): p. 1-7.
37. Heit, J.A., F.A. Spencer, and R.H. White, *The epidemiology of venous thromboembolism.* J Thromb Thrombolysis, 2016. **41**(1): p. 3-14.

38. Giordano, N.J., et al., *Epidemiology, Pathophysiology, Stratification, and Natural History of Pulmonary Embolism.* Techniques in Vascular and Interventional Radiology, 2017.
39. General, O.o.t.S., *The Surgeon General's call to action to prevent deep vein thrombosis and pulmonary embolism.* 2008.
40. Heit, J.A., et al., *Reasons for the persistent incidence of venous thromboembolism.* Thromb Haemost, 2017. **117**(2): p. 390-400.
41. Prandoni, P., *Venous and arterial thrombosis: two aspects of the same disease?* Clinical epidemiology, 2009. **1**: p. 1.
42. Ro, A., N. Kageyama, and T. Mukai, *Pathophysiology of venous thromboembolism with respect to the anatomical features of the deep veins of lower limbs: a review.* Annals of vascular diseases, 2017: p. ra. 17-00035.
43. Schwartz, T., et al., *Pulmonary embolism without deep venous thrombosis.* Annals of vascular surgery, 2012. **26**(7): p. 973-976.
44. Van Gent, J.-M., et al., *Pulmonary embolism without deep venous thrombosis: de novo or missed deep venous thrombosis?* Journal of Trauma and Acute Care Surgery, 2014. **76**(5): p. 1270-1274.
45. Bourget-Murray, J., et al., *Symptomatic bilateral pulmonary embolism without deep venous thrombosis in an adolescent following arthroscopic anterior cruciate ligament reconstruction: a case report and review of the literature.* Journal of medical case reports, 2018. **12**(1): p. 194.
46. Flanders, S.A., et al., *Hospital performance for pharmacologic venous thromboembolism prophylaxis and rate of venous thromboembolism : a cohort study.* JAMA Intern Med, 2014. **174**(10): p. 1577-84.
47. Chua, C.C., et al., *Retrospective evaluation of Venous Thromboembolism (VTE): are all transient provoking events the same?* Eur J Haematol, 2017.
48. Boddi, M. and A. Peris, *Deep Vein Thrombosis in Intensive Care.* Adv Exp Med Biol, 2017. **906**: p. 167-181.
49. Monti, M., et al., *[Venous thromboembolism in critically ill patients: analysis of the main age-related risk factors and definition of specific scores.].* Recenti Prog Med, 2016. **107**(9): p. 480-484.
50. Ramot, Y., A. Nyska, and G. Spectre, *Drug-induced thrombosis: an update.* Drug safety, 2013. **36**(8): p. 585-603.
51. Parkin, L., et al., *Antidepressants, Depression, and Venous Thromboembolism Risk: Large Prospective Study of UK Women.* J Am Heart Assoc, 2017. **6**(5).
52. Lai, S., et al., *Venous Thromboembolism Rates in Transferred Patients: A Cross-Sectional Study.* Journal of General Internal Medicine, 2017.
53. Puurunen, M.K., et al., *Epidemiology of venous thromboembolism in the Framingham Heart Study.* Thromb Res, 2016. **145**: p. 27-33.
54. Souto, J.C., et al., *Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia.* Am J Hum Genet, 2000. **67**(6): p. 1452-9.
55. Margaglione, M. and E. Grandone, *Population genetics of venous thromboembolism. A narrative review.* Thromb Haemost, 2011. **105**(2): p. 221-31.
56. Rajendran, P., et al., *The vascular endothelium and human diseases.* Int J Biol Sci, 2013. **9**(10): p. 1057-69.

57. Kelly, J., et al., *Dehydration and venous thromboembolism after acute stroke.* QJM: An International Journal of Medicine, 2004. **97**(5): p. 293-296.
58. Jönsson, A.K., et al., *Venous Thromboembolism During Treatment with Antipsychotics: A Review of Current Evidence.* CNS Drugs, 2018. **32**(1): p. 47-64.
59. Minet, C., et al., *Venous thromboembolism in the ICU: main characteristics, diagnosis and thromboprophylaxis.* Crit Care, 2015. **19**(1): p. 287.
60. Kearon, C., et al., *Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH.* Journal of Thrombosis and Haemostasis, 2016. **14**(7): p. 1480-1483.
61. Zöller, B., *Genetics of venous thromboembolism revised.* Blood, The Journal of the American Society of Hematology, 2019. **134**(19): p. 1568-1570.
62. Lindström, S., et al., *Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism.* Blood, 2019. **134**(19): p. 1645-1657.
63. Ahmad, A., et al., *Association between TLR9 rs5743836 polymorphism and risk of recurrent venous thromboembolism.* J Thromb Thrombolysis, 2017.
64. Alpert, J.S. and J.E. Dalen, *Epidemiology and natural history of venous thromboembolism.* Prog Cardiovasc Dis, 1994. **36**(6): p. 417-22.
65. Anderson, F.A., Jr., et al., *A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism. The Worcester DVT Study.* Arch Intern Med, 1991. **151**(5): p. 933-8.
66. Skrifvars, M.B., et al., *Venous thromboembolic events in critically ill traumatic brain injury patients.* Intensive Care Med, 2017. **43**(3): p. 419-428.
67. Ramzi, D.W. and K.V. Leeper, *DVT and pulmonary embolism: Part I. Diagnosis.* Am Fam Physician, 2004. **69**(12): p. 2829-36.
68. Guyatt, G.H., et al., *Executive summary: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines.* Chest, 2012. **141**(2 Suppl): p. 7s-47s.
69. Lyman, G.H., et al., *American Society of Clinical Oncology guideline: recommendations for venous thromboembolism prophylaxis and treatment in patients with cancer.* Journal of clinical oncology, 2007. **25**(34): p. 5490-5505.
70. Rogers, F.B., et al., *Determining venous thromboembolic risk assessment for patients with trauma: the Trauma Embolic Scoring System.* Journal of Trauma and Acute Care Surgery, 2012. **73**(2): p. 511-515.
71. Hamada, S.R., et al., *High level of venous thromboembolism in critically ill trauma patients despite early and well-driven thromboprophylaxis protocol.* Ann Intensive Care, 2017. **7**(1): p. 97.
72. Deutsch, G.B., et al., *Bleeding risk secondary to deep vein thrombosis prophylaxis in patients with lower gastrointestinal bleeding.* Journal of intensive care medicine, 2012. **27**(6): p. 379-383.
73. Russell, R.D. and M.H. Huo, *Apixaban and rivaroxaban decrease deep venous thrombosis but not other complications after total hip and total knee arthroplasty.* The Journal of arthroplasty, 2013. **28**(9): p. 1477-1481.
74. G, M., *Preventing hospital-associated venous thromboembolism: a guide for effective quality improvement, 2nd ed*. August 2016, AHRQ: Rockville, MD: .
75. Chen, L., et al., *Dynamic and Personalized Risk Forecast in Step-Down Units: Implications for Monitoring Paradigms.* Ann Am Thorac Soc, 2016.

76. Pinsky, M.R. and A. Dubrawski, *Gleaning Knowledge from Data in the Intensive Care Unit.* American Journal of Respiratory and Critical Care Medicine, 2014. **190**(6): p. 606-610.
77. Shmueli, G., *To explain or to predict?* Statistical science, 2010. **25**(3): p. 289-310.
78. Oquendo, M.A., et al., *Machine learning and data mining: strategies for hypothesis generation.* Molecular Psychiatry, 2012. **17**(10): p. 956-959.
79. Bose, E.L., et al., *Cardiorespiratory instability in monitored step-down unit patients: using cluster analysis to identify patterns of change.* J Clin Monit Comput, 2017.
80. Guillame-Bert, M., et al., *Learning temporal rules to forecast instability in continuously monitored patients.* J Am Med Inform Assoc, 2017. **24**(1): p. 47-53.
81. Harris, P.R., et al., *Prognostic value of heart rate turbulence for risk assessment in patients with unstable angina and non-ST elevation myocardial infarction.* Vasc Health Risk Manag, 2013. **9**: p. 465-73.
82. Pinsky, M.R., G. Clermont, and M. Hravnak, *Predicting cardiorespiratory instability.* Crit Care, 2016. **20**: p. 70.
83. Frisch, S.O., et al., *Abstract 14879: Using Predictive Machine Learning Modeling for the Nursing Triage of Acute Chest Pain at the Emergency Department.* Circulation, 2019. **140**(Suppl_1): p. A14879-A14879.
84. Xu, J., et al., *Review and evaluation of electronic health records-driven phenotype algorithm authoring tools for clinical and translational research.* J Am Med Inform Assoc, 2015. **22**(6): p. 1251-60.
85. Chen, Y., et al., *Applying active learning to high-throughput phenotyping algorithms for electronic health records data.* J Am Med Inform Assoc, 2013. **20**(e2): p. e253-9.
86. McPeek Hinz, E.R., L. Bastarache, and J.C. Denny, *A natural language processing algorithm to define a venous thromboembolism phenotype.* AMIA Annu Symp Proc, 2013. **2013**: p. 975-83.
87. Richesson, R.L., et al., *Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory.* Journal of the American Medical Informatics Association, 2013. **20**(e2): p. e226-e231.
88. Richesson, R.L., M.M. Smerek, and C. Blake Cameron, *A Framework to Support the Sharing and Reuse of Computable Phenotype Definitions Across Health Care Delivery and Clinical Research Applications.* eGEMs, 2016. **4**(3): p. 1232.
89. Richesson R, S.M., *Electronic Health Records-Based Phenotyping*, in *Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials*. 2017, NIH Health Care Systems Research Collaboratory: Bethesda, MD.
90. Liao, K.P., et al., *Development of phenotype algorithms using electronic medical records and incorporating natural language processing.* bmj, 2015. **350**: p. h1885.
91. Banda, J.M., et al., *Advances in electronic phenotyping: from rule-based definitions to machine learning models.* Annual review of biomedical data science, 2018. **1**: p. 53-68.
92. Yount, R.J., J.K. Vries, and C.D. Councill, *The medical archival system: an information retrieval system based on distributed parallel processing.* Information processing & management, 1991. **27**(4): p. 379-389.
93. Provost, F. and T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*. 2013: " O'Reilly Media, Inc.".

94. Alotaibi, G.S., et al., *The validity of ICD codes coupled with imaging procedure codes for identifying acute venous thromboembolism using administrative data.* Vascular Medicine, 2015. **20**(4): p. 364-368.
95. Pellathy, T., et al., *205: ACCURACY OF IDENTIFYING VENOUS THROMBOEMBOLISM BY ADMINISTRATIVE CODING COMPARED TO MANUAL REVIEW.* Critical Care Medicine, 2018. **46**(1): p. 85.
96. *The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*. 2007, U.S. Department of Health and Human Services.
97. Le Gal, G. and M. Righini, *Controversies in the diagnosis of venous thromboembolism.* Journal of Thrombosis and Haemostasis, 2015. **13**(S1).
98. Pellathy, T., et al. *Prevalence of Venous Thromboembolism (VTE) in an Adult Step-Down Unit Population: A Proof-of-Concept Feasibility Study for Machine Learning Predictive Model Development*. in *NURSING RESEARCH*. 2018. LIPPINCOTT WILLIAMS & WILKINS TWO COMMERCE SQ, 2001 MARKET ST, PHILADELPHIA ….
99. Pellathy, T., et al., *58: IDENTIFYING TIME INTERVAL FEATURES PREDICTIVE OF HOSPITAL-ACQUIRED VENOUS THROMBOEMBOLISM.* Critical Care Medicine, 2019. **47**(1): p. 29.
100. Figueroa, R.L., et al., *Predicting sample size required for classification performance.* BMC Medical Informatics and Decision Making, 2012. **12**: p. 8-8.
101. Kantardzic, M., *Data mining: concepts, models, methods, and algorithms*. 2011: John Wiley & Sons.
102. Spyropoulos, A.C., et al., *Rates of venous thromboembolism occurrence in medical patients among the insured population.* Thromb Haemost, 2009. **102**(5): p. 951-7.
103. Moheimani, F. and D.E. Jackson, *Venous thromboembolism: classification, risk factors, diagnosis, and management.* ISRN Hematol, 2011. **2011**: p. 124610.
104. Faul, F., et al., *G* Power Version 3.1. 7 [computer software] Uiversität Kiel; Germany: 2013*. 2013.
105. Benchimol, E.I., et al., *The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement.* PLoS medicine, 2015. **12**(10): p. e1001885.
106. Von Elm, E., et al., *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies.* International Journal of Surgery, 2014. **12**(12): p. 1495-1499.
107. Denaxas, S., et al., *Methods for enhancing the reproducibility of biomedical research findings using electronic health records.* BioData Mining, 2017. **10**(1): p. 31.
108. *LOINC: The international standard for identifying health measurements, observations, and documents.* 2017 [cited 2017 October 22, 2017]; Available from: https://loinc.org.
109. Musen, M.A., et al., *The center for expanded data annotation and retrieval.* Journal of the American Medical Informatics Association : JAMIA, 2015. **22**(6): p. 1148-1152.
110. Kroegel, C. and A. Reissig, *Principle mechanisms underlying venous thromboembolism: epidemiology, risk factors, pathophysiology and pathogenesis.* Respiration, 2003. **70**(1): p. 7-30.
111. Posch, F., et al., *Dynamic assessment of venous thromboembolism risk in patients with cancer by longitudinal D-Dimer analysis: A prospective study.* Journal of Thrombosis and Haemostasis, 2020.
112. Streiff, M.B., et al., *CDC Grand Rounds: preventing hospital-associated venous thromboembolism.* MMWR Morb Mortal Wkly Rep, 2014. **63**(9): p. 190-3.

113. Tseytlin, E., et al., *NOBLE–Flexible concept recognition for large-scale biomedical natural language processing.* BMC bioinformatics, 2016. **17**(1): p. 32.
114. Tabachnick, B. and L. Fidell, *Using multivariate statistics . Harlow*. 2014, United Kingdom: Pearson.
115. Belsley, D.A., *A guide to using the collinearity diagnostics.* Computer Science in Economics and Management, 1991. **4**(1): p. 33-50.
116. Chawla, N.V., et al., *SMOTE: synthetic minority over-sampling technique.* Journal of artificial intelligence research, 2002. **16**: p. 321-357.
117. Goldhaber, S.Z. and H. Bounameaux, *Pulmonary embolism and deep vein thrombosis.* The Lancet, 2012. **379**(9828): p. 1835-1846.
118. Darzi, A., et al., *Risk models for VTE and bleeding in medical inpatients: Systematic identification and expert assessment.* 2020.
119. Maynard, G., *Preventing hospital: associated venous thromboembolism: a guide for effective quality improvement [Internet]. Rockville: AHRQ; 2016 [cited 2017 Oct 14]*.
120. Bose, E., et al., *Risk for Cardiorespiratory Instability Following Transfer to a Monitored Step-Down Unit.* Respir Care, 2017. **62**(4): p. 415-422.
121. Team, R.C., *R: A Language and Environment for Statistical Computing*, R.F.f.S. Computing, Editor. 2020: Vienna, Austria.
122. van Buuren, S., Groothius-Oudshoorn, K., *Multivariate Imputation by Chained Equations in R.* Journal of Statistical Software, 2011. **45**(3): p. 1-67.
123. Leisman, D.E., et al., *Development and reporting of prediction models: guidance for authors from editors of respiratory, sleep, and critical care journals.* Critical care medicine, 2020. **48**(5): p. 623.
124. Wang, X., et al., *Comparing different venous thromboembolism risk assessment machine learning models in Chinese patients.* Journal of Evaluation in Clinical Practice, 2020. **26**(1): p. 26-34.
125. Caprini, J., et al. *Clinical assessment of venous thromboembolic risk in surgical patients*. in *Seminars in thrombosis and hemostasis*. 1991.
126. Greene, M.T., et al., *Validation of risk assessment models of venous thromboembolism in hospitalized medical patients.* The American Journal of Medicine, 2016. **129**(9): p. 1001. e9-1001. e18.
127. Grant, P.J., et al., *Assessing the Caprini Score for Risk Assessment of Venous Thromboembolism in Hospitalized Medical Patients.* Am J Med, 2016. **129**(5): p. 528-35.
128. Gerotziafas, G.T., et al., *Updated clinical models for VTE prediction in hospitalized medical patients.* Thromb Res, 2018. **164 Suppl 1**: p. S62-s69.
129. Blondon, M., et al., *External validation of the simplified Geneva risk assessment model for hospital-associated venous thromboembolism in the Padua cohort.* Journal of Thrombosis and Haemostasis, 2020. **18**(3): p. 676-680.
130. Bahl, V., et al., *A validation study of a retrospective venous thromboembolism risk scoring method.* Annals of surgery, 2010. **251**(2): p. 344-350.
131. Ferroni, P., et al., *Risk assessment for venous thromboembolism in chemotherapy-treated ambulatory cancer patients: a machine learning approach.* Medical Decision Making, 2017. **37**(2): p. 234-242.
132. James, S., et al., *Novel Algorithms to Predict the Occurrence of In-Hospital Venous Thromboembolism: Machine Learning Classifiers Developed From the 2012 National Inpatient Sample.* Chest, 2015. **148**(4): p. 492A.

133. Kawaler, E., et al. *Learning to predict post-hospitalization VTE risk from EHR data*. in *AMIA annual symposium proceedings*. 2012. American Medical Informatics Association.