

Random forests could help to evaluate interventions to improve thromboprophylaxis regimens for inpatients

Automated identification of hospital-acquired venous thromboembolism

Patrick E. Beeler^{a,b,c}, Qoua L. He^a, Adam Wright^{a,b}, David W. Bates^{a,b}

^a Division of General Internal Medicine & Primary Care, Brigham and Women's Hospital, Boston, MA, USA; ^b Harvard Medical School, Boston, MA, USA; ^c Research Center for Medical Informatics, University Hospital Zurich, Switzerland

Introduction

Venous thromboembolism (VTE) as a hospital-acquired condition – i.e., not ‘present on admission’ – is a potentially preventable complication. A decrease of hospital-acquired VTE events indicates success of efforts to prevent VTE in hospitalized patients. However, so far, costly chart reviews have been needed to identify patients with hospital-acquired VTE. We investigated whether electronic health record data such as medication orders and their temporal relations allow for differentiating between hospital-acquired VTE and VTE present on admission. Therefore, we modeled a tree and two random forests and evaluated the automated classification of hospital-acquired VTE.

Methods

All inpatients with a length of stay of ≥ 24 h, discharged from the Brigham and Women's Hospital, a large tertiary care hospital in Boston, MA, between January 2009 and April 2014 were searched for ICD-9 diagnosis codes of acute venous thrombosis or pulmonary embolism. Patients were included who had VTE in the admitting diagnosis field – defined as VTE present on admission – or in one of up to 50 discharge diagnoses. Of those, only patients who received heparin, dalteparin, enoxaparin, alteplase, rivaroxaban or fondaparinux were considered, and the time from admission to the first order was calculated for each drug. Additionally included predictors were: dose information, demographics (age, gender, race, language), length of stay, admission service, discharge service, transfer destination of the patient after discharge, and whether the patient was alive or died during the hospitalization or within 30 days after discharge. A single tree and two random forests (each with 5000 trees) were generated to analyze the predictors and to assess the predictive power of the chosen approach. Since medication orders are electronically available in real time, such prospective predictors may have implications for clinical decision support. Therefore, prospective predictors (i.e., demographics, admission service, time to order a drug, route and dose information for each drug) were separately analyzed in the first random forest.

Half of the data served as calibration set, half as validation set. Statistical computing was performed using the software R version 3.1.0 (R Foundation for Statistical Computing, Vienna, Austria).

Results

A total of 5374 patient stays featured a VTE diagnosis with a defined drug order. If VTE was present on admission ($n = 1262$; 23.5%), the median time to order one of the aforementioned drugs was 2.5 h (interquartile range [IQR] 1.3–5.0 h). Among hospital-acquired VTE cases without an admitting diagnosis of VTE ($n = 4112$; 76.5%), the median time to order the drug was 4.2 h (IQR 1.7–18.2 h). Unsurprisingly, a single tree – after cross-validation and pruning – identified the time from admission to the ordering of intravenous (IV) heparin as the most significant predictor (fig. 1). This tree's validation resulted in an accuracy of 78.8% and a positive predictive value (PPV) of 83.3% for the classification of hospital-acquired VTE.

The first validated random forest used predictors that are available in real time: the forest had an accuracy of 79.7% and a PPV of 85.3% for the classification of hospital-acquired VTE. The second validated random forest considered all variables and resulted in an accuracy of 81.7% and a PPV of 87.8% (the importance of the variables is shown in fig. 2).

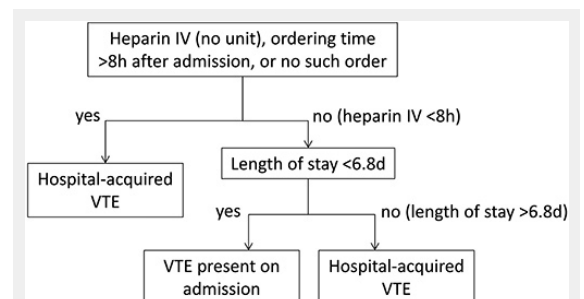


Figure 1:

Cross-validated and pruned tree (all variables considered).

Discussion

We modeled a tree and two random forests using structured data predictors to differentiate between hospital acquired VTE and VTE present on admission. Our validated tree (fig. 1), considering the first order for IV heparin and the

length of stay, could immediately be implemented as a first step to identifying patients with hospital-acquired VTE patients. However, the random forests performed better, even when exclusively prospective predictors were used, and such real time models may have implications for clinical decision support tools. In conclusion, our random forests could help to evaluate interventions to improve thromboprophylaxis regimens for inpatients, where costly chart reviews are needed to differentiate between VTE present on admission and potentially preventable complications.

Disclosure statement

No potential conflict of interest relevant to this article was reported.

Correspondence:

Dr. med. Patrick Emanuel Beeler

Centre on Aging and Mobility

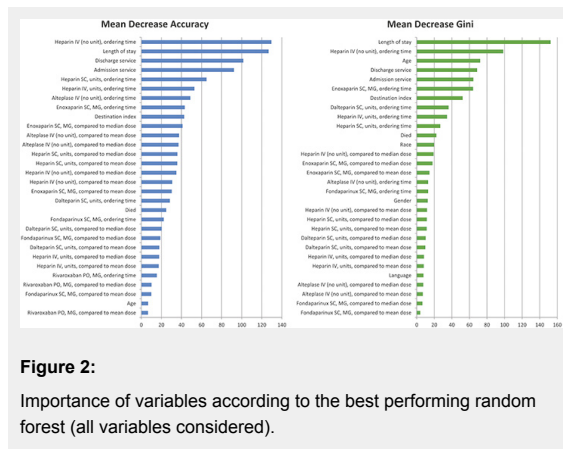
Dept. of Geriatrics and Aging Research

University Hospital Zürich and University of Zürich

Tièchestrassè 99

CH-8037 Zürich

[patrick.beeler\[at\]jusz.ch](mailto:patrick.beeler[at]jusz.ch)



Figures (large format)

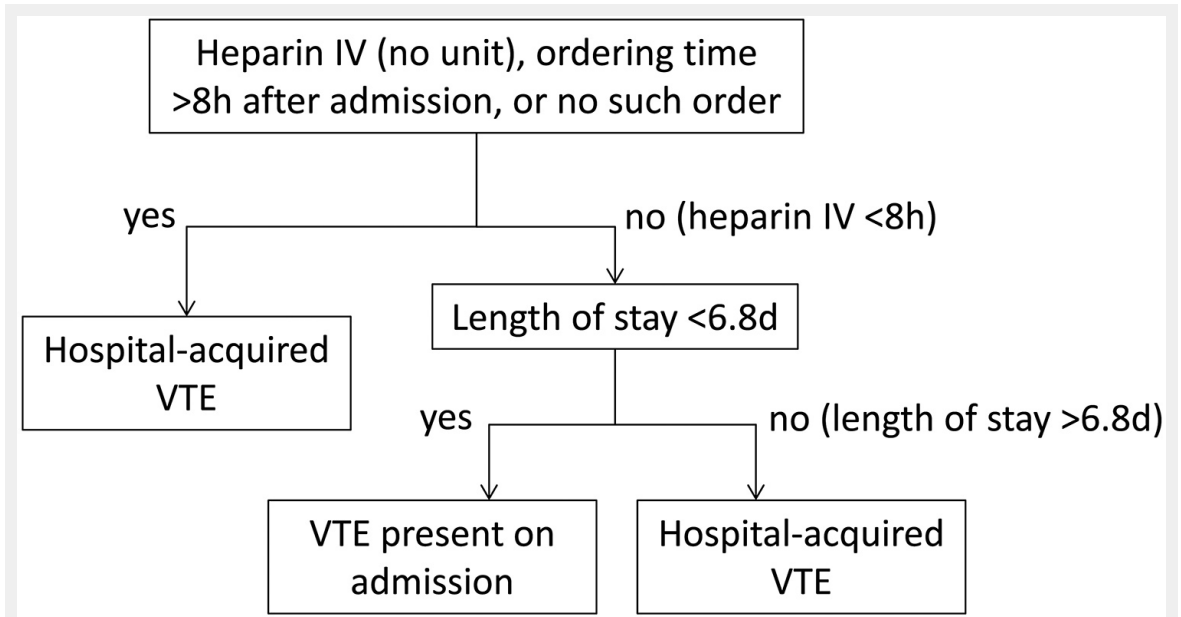


Figure 1: Cross-validated and pruned tree (all variables considered).

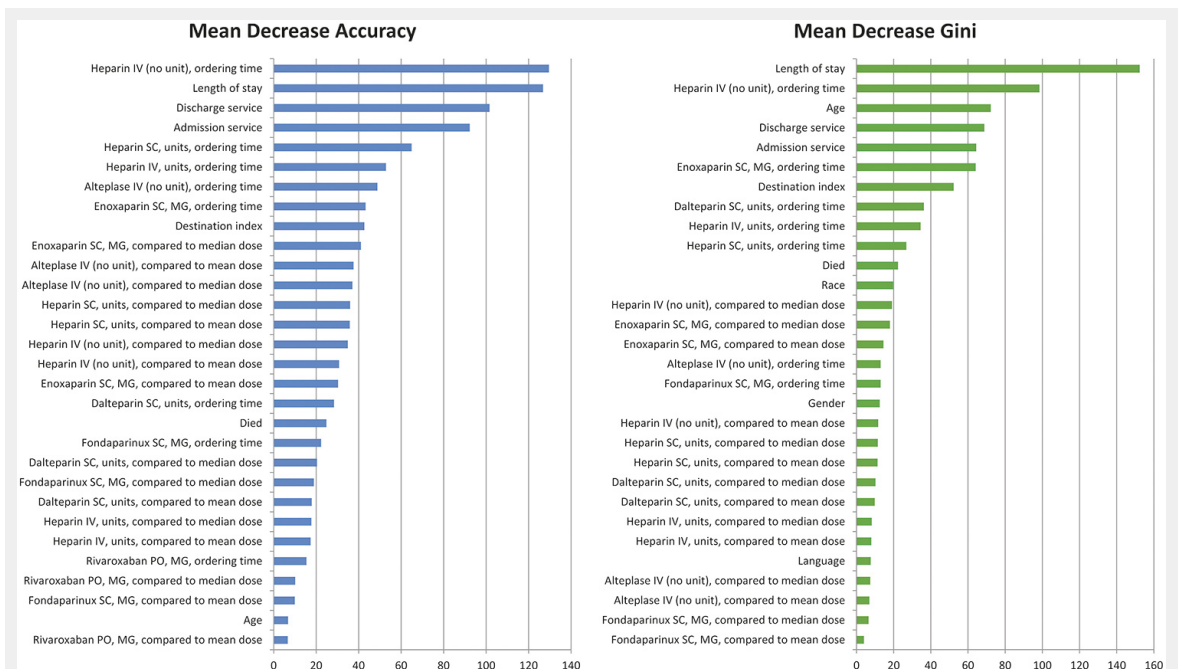


Figure 2: Importance of variables according to the best performing random forest (all variables considered).