



Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2016 November ; 25(11): 1228–1235. doi:10.1002/pds.4069.

Predicting prolonged dose titration in patients starting warfarin

Brian S. Finkelman, PhD^{1,5}, Benjamin French, PhD¹, Luanne Bershaw, LPN¹, Colleen M. Brensinger, MS¹, Michael B. Streiff, MD², Andrew E. Epstein, MD^{3,4}, and Stephen E. Kimmel, MD, MSCE^{1,3,5}

¹Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

²Department of Medicine, Hematology Division, Johns Hopkins University School of Medicine, Baltimore, MD

³Department of Medicine, Cardiovascular Division, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

⁴Department of Medicine, Corporal Michael J. Crescenz Veterans Affairs Medical Center, Philadelphia, PA

⁵Center for Therapeutic Effectiveness Research, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Abstract

Purpose—Patients initiating warfarin therapy generally experience a dose-titration period of weeks to months, during which time they are at higher risk of both thromboembolic and bleeding events. Accurate prediction of prolonged dose titration could help clinicians determine which patients might be better treated by alternative anticoagulants that, while more costly, do not require dose titration.

Methods—A prediction model was derived in a prospective cohort of patients starting warfarin (N = 390), using Cox regression, and validated in an external cohort (N = 663) from a later time period. Prolonged dose titration was defined as a dose-titration period >12 weeks. Predictor variables were selected using a modified best subsets algorithm, using leave-one-out cross-validation (LOOCV) to reduce overfitting.

Results—The final model had five variables: warfarin indication, insurance status, number of doctor's visits in the previous year, smoking status, and heart failure. The area under the ROC curve (AUC) in the derivation cohort was 0.66 (95% CI 0.60, 0.74) using LOOCV, but only 0.59

Corresponding address: Stephen E. Kimmel, 923 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, *Contact information:* Office: 215-898-1740, Fax: 215-573-3106, stevek@mail.med.upenn.edu.

Conflicts of Interest:

None of the other authors have any conflicts of interest to disclose.

Prior Presentations:

None.

Ethics statement

The study was approved by the Institutional Review Boards at the University of Pennsylvania, the Corporal Michael J. Crescenz Veterans Affairs Medical Center, Hershey Medical Center, and Johns Hopkins University. All study participants provided written informed consent.

(95% CI 0.54, 0.64) in the external validation cohort, and varied across clinics. Including genetic factors in the model did not improve the AUC (0.59; 95% CI 0.54, 0.65). Relative utility curves indicated that the model was unlikely to provide a clinically meaningful benefit compared to no prediction.

Conclusions—Our results suggest that prolonged dose titration cannot be accurately predicted in warfarin patients using traditional clinical, social, and genetic predictors, and that accurate prediction will need to accommodate heterogeneities across clinical sites and over time.

Keywords

warfarin; treatment effectiveness; risk assessment; decision analysis; validity (epidemiology); pharmacogenetics

Introduction

Because of the substantial population variability in warfarin dose requirements, patients starting warfarin therapy often experience a lengthy dose-titration period of weeks to months. During this period, patients are at particularly high risk of both bleeding and thromboembolic complications from improper anticoagulation levels.^{1,2} Prolonged dose titration also increases patient burden by adding more frequent monitoring of the international normalized ratio (INR), which can lead to a reduced quality of life and higher rates of discontinuation of a highly efficacious therapy.^{3–5} Given the availability of less burdensome but more expensive alternative oral anticoagulants,⁶ accurate identification of patients most likely to experience a prolonged dose-titration period on warfarin could potentially help clinicians decide when to use warfarin versus one of the alternative agents.

In this study, we derived a prediction model for whether patients initiating warfarin experienced a prolonged dose-titration period using a prospective cohort of patients starting warfarin therapy at specialty anticoagulation clinics (N = 390). We then validated this model in a similar, but independent, external cohort of patients initiating warfarin (N = 663). Finally, we tested whether inclusion of genetic factors—specifically, *CYP2C9* and *VKORC1* variants—improved model accuracy.

Methods

Derivation cohort

We derived the model using the IN-RANGE cohort, a large prospective cohort of warfarin initiation that has been used to study the clinical and genetic predictors of warfarin maintenance dose and adherence.^{7–15} Participants were from specialty anticoagulation clinics at the Hospital of the University of Pennsylvania (HUP), the Corporal Michael J. Crescenz Veterans Affairs Medical Center (CMCVAMC), and Hershey Medical Center. Institutional review board (IRB) approval was obtained at all three sites, and all study participants provided written informed consent. Exclusion criteria were: age <21 years; being unwilling or unable to provide consent; having an abnormal INR prior to starting anticoagulation therapy; and having antiphospholipid antibodies. All participants in the

original IN-RANGE cohort (N = 390), enrolled between April 2002 and February 2006, were included in the derivation cohort for the current study.

Validation cohort

The cohort used for model validation was the IN-RANGE2 cohort, which was designed as a follow-up cohort to the original IN-RANGE cohort, with similar data-collection methods. Participants were recruited from specialty anticoagulation clinics at HUP, CPCVAMC, and Johns Hopkins University (JHU). IRB approval was obtained at all three sites, and all study participants provided written informed consent. Exclusion criteria were the same as the original IN-RANGE cohort, except individuals who were neither Caucasian nor African American (about 3% of the original cohort) were excluded, and individuals with antiphospholipid antibodies were no longer excluded from the IN-RANGE2 study. Participants were enrolled between October 2009 and August 2013. All participants with available follow-up data (N = 663) were included in the validation cohort for the current study.

Primary outcome

The primary outcome was a prolonged dose-titration phase, which was defined as achieving maintenance dose in >12 weeks of attempted warfarin therapy. Twelve weeks is a clinically meaningful cut-off because the first 3 months of warfarin therapy have been shown to be especially high risk for patients,¹ and some warfarin indications, such as venous thromboembolism with transient risk factors, often only require a 3-month course of therapy.¹⁶ Additionally, we chose a dichotomous outcome to make it easier for clinicians to incorporate model predictions into their decision-making process. Maintenance dose achievement was defined as having two consecutive INRs within the therapeutic range, at the same warfarin dose, at least one week apart. Using this definition allowed the outcome to be defined the same across both the derivation and validation cohorts. The time of maintenance dose achievement was taken as the number of days from warfarin initiation to the first maintenance dose-defining visit in days. Achieving maintenance dose within 4 and 8 weeks were also considered as outcomes in secondary analyses. Note that, because the derivation and validation cohorts were both observational, warfarin dose titration was performed at the discretion of each patient's individual practitioner; however, dosing within clinics was standardized and reflects typical practice at anticoagulation clinics.

Candidate predictors

A total of 28 baseline social and clinical factors were considered for inclusion in the prediction model, shown in Supplementary Table 1. Some of these candidate predictors have been previously associated with time to maintenance dose,¹⁷ and most have been associated with other warfarin-related outcomes, such as warfarin maintenance dose requirement,^{18,19} poor warfarin adherence,^{4,11,15} discontinuation of warfarin,^{5,20,21} percent time in therapeutic range,^{10,22,23} and risk of bleeding events.^{24–27} Additionally, we were interested in whether inclusion of genetic factors could improve model accuracy. For this analysis, we added genetic variants in *CYP2C9* (rs1799853 and rs1057910) and *VKORC1* (rs9923231), specified in a binary fashion as having at least one variant in the given gene, to the model. These variants were chosen because they have most consistently demonstrated a large

association with warfarin maintenance dose, are used in the major pharmacogenetic dosing algorithms, and were available in both derivation and validation cohort.^{18,19,28}

Model-building strategy

Because approximately 11% of the derivation cohort was censored prior to 12 weeks of attempted warfarin therapy, we used a Cox regression model with time from initiation of warfarin to the achievement of maintenance dose or censoring as the outcome. Variable selection was conducted using a modified best subsets algorithm. This algorithm, described in detail in the Supplementary Methods section, was designed to optimize model discrimination, or how well the model distinguished between those with a prolonged versus non-prolonged dose-titration phase. We used the area under the time-dependent ROC curve (AUC) to measure model discrimination.²⁹ Leave-one-out cross-validation (LOOCV)—in which the model is derived in all individuals in the dataset except one, who is then used to test the model, repeated over all individuals in the dataset—was used when estimating the AUC, in order to avoid selecting a model because of overfitting. Thus, our algorithm selected the combination of candidate variables with the best AUC, as estimated by LOOCV. To further reduce the chance of overfitting, we limited the models examined to no more than 10 variables. The variables selected for the prediction model were then inspected graphically to ensure proper functional form, and all coefficients were examined to ensure that the direction of effect reported by the model was consistent with the available literature. A linear shrinkage factor, estimated as the mean slope of the observed outcome regressed on the linear predictors from models fit in 1,000 bootstrap samples, was then applied to all coefficients in this final model to improve model calibration without sacrificing model discrimination.³⁰

Model assessment and validation

The performance of the final prediction model was assessed in a separate validation cohort, described above. The primary metric of model performance was the AUC. We also tested whether adding genetic predictors improved model accuracy, based on the AUC and the integrated discrimination improvement (IDI).³¹ We also assessed the calibration of the prediction model using calibration plots, and we examined the clinical utility of the prediction model using plots of the relative utility of the model versus the risk threshold,^{32,33} which are described in more detail in the Supplementary Methods. Finally, we conducted several *post-hoc* analyses, including assessing site-specific differences in model performance, deriving and validating the model in the two anticoagulation clinics that were common across both cohorts (HUP and CMCVAMC), and implementing the above variable selection algorithm in the validation cohort. Confidence intervals for all estimates were generated using the 2.5th and 97.5th percentiles of estimates in 1,000 bootstrap replications. All analyses were completed using R 3.1.0.³⁴

Results

The characteristics of the derivation and validation cohorts are shown in Table 1. The overall incidence of prolonged dose-titration was 30% in the derivation cohort and 38% in the validation cohort. The variable-selection algorithm found that the model with the best

discrimination, as estimated by LOOCV, contained the following five variables: warfarin indication, insurance status, number of doctor's visits in the previous year, current smoking status, and history of congestive heart failure. The proportional hazards assumption was not violated for this model ($P=0.83$), as tested using the method of Grambsch and Therneau.³⁵ The model had modest accuracy to discriminate between patients who did and did not experience prolonged dose titration in the derivation cohort (LOOCV AUC, 0.66; 95% CI 0.60, 0.74) (Figure 1). A comparison of this model to the other top-performing models with different numbers of predictor variables, as measured with and without cross-validation, suggested that using cross-validation successfully avoided complex models that were more accurate merely because of having extra degrees of freedom (Supplementary Figure 1). Furthermore, the variables selected were stable, with the same set of five important variables included in all of the best models with five or more variables (Supplementary Table 2). In order to account for overfitting, we estimated a shrinkage factor based on 1,000 bootstrap replications at 0.82, indicating a moderate degree of overfitting by the original model. Coefficients from the final prediction model, after applying the linear shrinkage factor, are shown in Table 2.

When evaluated in the validation cohort, the AUC of the prediction model at 12 weeks was 0.59 (95% CI 0.54, 0.64) (Figure 1). In secondary analyses, the AUC of the model at 8 weeks was 0.57 (95% CI 0.53, 0.62) and at 4 weeks was 0.57 (95% CI 0.52, 0.62). The calibration of the main model was examined by comparing observed to predicted probabilities across risk deciles (Supplementary Figure 2); the Hosmer-Lemeshow test for goodness of fit did not demonstrate poor model calibration ($P=0.75$).

Addition of genetic variants in *CYP2C9* and *VKORC1* to the model did not improve model accuracy ($P=0.98$), with the AUC unchanged at 0.59 (95% CI 0.54, 0.65) (Supplementary Figure 3). The IDI from adding genetic factors to the model was estimated as 0.010 (95% CI 0.002, 0.018), which is equivalent to a 7% increase in model discrimination over the model without genetic factors.

Table 3 shows the sensitivity, specificity, and positive and negative predictive values for various risk thresholds of the prediction model. Similarly, Figure 2 shows the relationship between the positive and negative predictive values and the proportion that are classified as positive across the full range of risk thresholds. The relative utility of the model across the full range of risk thresholds is shown in Figure 3. The relative utility had a maximum value of 9.6% and negative values for the risk threshold range of 48% to 62%. Relative utility curves for the models with and without genetic factors are shown in Supplementary Figure 4.

Differences in the characteristics of the derivation and validation cohorts at HUP and CPCVAMC are shown in Supplementary Table 3. In *post-hoc* analyses, the AUC at 12 weeks was 0.59 (95% CI 0.51, 0.67) at HUP, 0.55 (95% CI 0.44, 0.62) at CPCVAMC, and 0.61 (95% CI 0.53, 0.70) at JHU. The observed frequency of prolonged dose titration was 32%, 34%, and 47% at HUP, CPCVAMC, and JHU, respectively, while the predicted probabilities of the outcome at these sites were 37%, 39%, and 38%. Deriving and validating the model in the same clinics, HUP and CPCVAMC, showed no improvement in model

performance (AUC = 0.58; 95% CI 0.51, 0.64). Finally, implementing the variable selection algorithm in the validation cohort led to the selection of different variables (Supplementary Table 4) and similar model performance on cross-validation (LOOCV AUC = 0.62; 95% CI 0.58, 0.69).

Discussion

In this study, we developed a model to predict whether a patient starting warfarin would experience a prolonged dose-titration phase, and tested this model in an external validation cohort. The final model performed modestly in the derivation cohort, with an LOOCV AUC at 12 weeks of 0.66 (95% CI 0.60, 0.74). However, the model performed worse when validated externally, with an AUC at 12 weeks of only 0.59 (95% CI 0.54, 0.64), confirming the generally recognized importance of external validation for prediction models.³⁶ Model performance did not improve for the secondary outcomes of reaching maintenance dose within 4 and 8 weeks, indicating that the model's poor discrimination was not unique to a specific cut-off.

The addition of genetic variants in *CYP2C9* or *VKORC1* did not improve model performance, with no improvement in the AUC observed ($P=0.98$). Similarly, the IDI was also poor at 0.010 (95% CI 0.002, 0.018). Although the IDI was statistically significant, it is known to have problems with inflated type I error, especially as it approaches zero;^{37,38} this finding should thus be viewed skeptically in the context of our overall results. Furthermore, the lack of improvement from adding genetic factors is consistent with recent clinical trial evidence showing that inclusion of genetic factors in dose prediction models did not lead to significant improvement in clinical outcomes, such as percent time in therapeutic range, compared with purely clinical dosing algorithms.²⁸

Attempts to quantify the clinical impact of the prediction model were consistent with our primary results. While the negative predictive value of the model for the lowest range of predicted values (<20% probability of prolonged dose titration) was reasonably good at 0.87, only 2% of patients in the validation cohort actually fall into this category (Table 3). The marked drop-off in performance at more commonly observed cut-offs may result from incorrectly ranking individuals in the middle of the probability distribution, which can be seen when plotting the observed vs predicted probabilities by risk decile (Supplementary Figure 2). As shown in Figure 3, the relative utility of the current model is limited, with a maximum value of 9.6% near the prevalence of the outcome and negative values for risk thresholds from 48% to 62%.

The failure of the model to validate is likely related to substantial differences between the two cohorts, as shown in Table 1. Compared with the derivation cohort, the validation cohort was younger, more African American, more obese, more under-insured, and more disabled, among other differences. These differences probably contributed to the higher incidence of prolonged dose titration in the validation cohort (38%) than the derivation cohort (30%). Part of these differences might reflect differing patient populations across sites; for example, the anticoagulation clinic at Johns Hopkins draws from a much more urban African American population than the clinic at Hershey Medical Center. These site-specific differences may

have translated into variable model performance, with the AUC at 12 weeks ranging from 0.55 to 0.61 across sites.

However, there are also substantial differences within the same sites between the two cohorts, including the proportion of individuals who are African American, the prevalence of hypertension, the prevalence of different warfarin indications, and the proportion of individuals on disability (Supplementary Table 3). These differences could be attributed to random fluctuations, changes in alternative anticoagulant use over time, or changes in recruitment strategies or participation rates between the two studies. For instance, individuals on disability may have had a more difficult time getting to anticoagulation clinics to adjust their warfarin dose; thus, an increase in the proportion of these individuals could result in an increase in the prevalence of prolonged dose titration that would not be accommodated by the prediction model, as the model does not include this predictor variable. The importance of these differences within the same sites over time is confirmed by the fact that there was no improvement in model performance when the model was derived and validated in the same clinical sites (AUC = 0.58; 95% CI 0.51, 0.64).

Taken in total, our *post-hoc* analyses suggest that important predictors of prolonged dose titration might vary across clinical sites and over time. For instance, implementing the same variable selection algorithm yielded a different set of important predictors, with different model coefficients, in the validation cohort (Supplementary Table 4) than in the derivation cohort (Table 2). Furthermore, as the broader differences among sites where patients receive warfarin in the clinical community would be expected to be much larger than the differences among the anticoagulation clinics in our derivation and validation cohorts, the important predictors of prolonged dose titration could be reasonably expected to show greater variation in clinical practice than reported here. Thus, to be able to accurately predict prolonged dose titration, future models will likely need to be able to accommodate such variation across sites and over time, using more advanced methods such as dynamic prediction modeling.^{39–41}

There are several strengths of our study. Most importantly, we utilized an external validation cohort, which is considered the gold standard for assessing the ability of a prediction model to make out-of-sample predictions.⁴² We also used a rigorous model-building strategy, incorporating cross-validation and coefficient shrinkage to help reduce the effects of overfitting. Finally, we directly assessed the clinical usefulness of our prediction model using positive and negative predictive values, as well as the more recent metric of relative utility.

However, there were also limitations to our study. Our prediction model was limited to variables that were available in both the derivation and validation cohorts. While the candidate variables included most of the traditional clinical and social factors that have been previously associated with warfarin-related outcomes, it is possible that other unknown or unmeasured factors, especially those related to health behaviors or health care access, might better predict prolonged dose titration. For instance, dietary vitamin-K consumption has been associated with anticoagulation control;^{43,44} thus, it is possible that measurement of dietary vitamin K could better predict prolonged dose titration among individuals and across

different clinical sites. We were also limited in the genotype data available, although the variants that were used are standard in warfarin dose prediction models. Because our cohorts included patients from specialty anticoagulation clinics, our results may not generalize to patients starting warfarin in other clinical settings.

In conclusion, our prediction model for prolonged dose titration in warfarin patients is unlikely to be useful in clinical practice. Furthermore, we suspect that prolonged dose titration will be difficult to predict using traditional clinical or genetic risk factors because the relationships between the predictors and the outcomes are likely to vary substantially across clinical sites and over time. Future models for predicting prolonged dose titration will need to use methods that can accommodate this variation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding source:

This work was funded by NIH NHLBI grant 5F30HL115992 to B.S.F. and NIH NHLBI grant 5R01HL066176 to S.E.K.

S.E.K. has consulted for several pharmaceutical companies, all unrelated to warfarin.

We would like to thank David Margolis, Scott Kasner, and Michael Levy for their helpful comments on the analyses and manuscript for this study.

References

1. Fihn SD, McDonnell M, Martin D, et al. Risk factors for complications of chronic anticoagulation. A multicenter study. Warfarin Optimized Outpatient Follow-up Study Group. *Ann Intern Med.* 1993; 118:511–520. [PubMed: 8280198]
2. Hylek EM, Skates SJ, Sheehan MA, Singer DE. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. *N Engl J Med.* 1996; 335:540–546. [PubMed: 8678931]
3. Dantas GC, Thompson BV, Manson JA, Tracy CS, Upshur REG. Patients' perspectives on taking warfarin: qualitative study in family practice. *BMC Fam Pract.* 2004; 5:15. [PubMed: 15268764]
4. Arnsten JH, Gelfand JM, Singer DE. Determinants of compliance with anticoagulation: A case-control study. *Am J Med.* 1997; 103:11–17. [PubMed: 9236480]
5. Fang MC, Go AS, Chang Y, et al. Warfarin discontinuation after starting warfarin for atrial fibrillation. *Circ Cardiovasc Qual Outcomes.* 2010; 3:624–631. [PubMed: 20959565]
6. Avorn J. The relative cost-effectiveness of anticoagulants: obvious, except for the cost and the effectiveness. *Circulation.* 2011; 123:2519–2521. [PubMed: 21606400]
7. Schelleman H, Brensinger CM, Chen J, Finkelman BS, Rieder MJ, Kimmel SE. New genetic variant that might improve warfarin dose prediction in African Americans. *Br J Clin Pharmacol.* 2010; 70:393–399. [PubMed: 20716240]
8. Schelleman H, Chen J, Chen Z, et al. Dosing algorithms to predict warfarin maintenance dose in Caucasians and African Americans. *Clin Pharmacol Ther.* 2008; 84:332–339. [PubMed: 18596683]
9. Kimmel SE, Christie J, Kealey C, et al. Apolipoprotein E genotype and warfarin dosing among Caucasians and African Americans. *Pharmacogenomics J.* 2008; 8:53–60. [PubMed: 17325732]

10. Kimmel SE, Chen Z, Price M, et al. The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch Intern Med.* 2007; 167:229–235. [PubMed: 17296877]
11. Platt AB, Localio AR, Brensinger CM, et al. Can we predict daily adherence to warfarin?: Results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Chest.* 2010; 137:883–889. [PubMed: 19903973]
12. Parker CS, Chen Z, Price M, et al. Adherence to warfarin assessed by electronic pill caps, clinician assessment, and patient reports: results from the IN-RANGE study. *J Gen Intern Med.* 2007; 22:1254–1259. [PubMed: 17587092]
13. Schelleman H, Chen Z, Kealey C, et al. Warfarin response and vitamin K epoxide reductase complex 1 in African Americans and Caucasians. *Clin Pharmacol Ther.* 2007; 81:742–747. [PubMed: 17329985]
14. Kealey C, Chen Z, Christie J, et al. Warfarin and cytochrome P450 2C9 genotype: possible ethnic variation in warfarin sensitivity. *Pharmacogenomics.* 2007; 8:217–225. [PubMed: 17324110]
15. Platt AB, Localio AR, Brensinger CM, et al. Risk factors for nonadherence to warfarin: results from the IN-RANGE study. *Pharmacoepidemiol Drug Saf.* 2008; 17:853–860. [PubMed: 18271059]
16. Agnelli G, Becattini C. Treatment of DVT: how long is enough and how do you predict recurrence. *J Thromb Thrombolysis.* 2008; 25:37–44. [PubMed: 17906973]
17. Finkelman BS, French B, Bershaw L, Kimmel SE. Factors affecting time to maintenance dose in patients initiating warfarin. *Pharmacoepidemiol Drug Saf.* 2015; 24:228–236. [PubMed: 25504915]
18. Gage BF, Eby C, Johnson JA, et al. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin Pharmacol Ther.* 2008; 84:326–331. [PubMed: 18305455]
19. Klein TE, Altman RB, Eriksson N, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med.* 2009; 360:753–764. [PubMed: 19228618]
20. Bushnell CD, Olson DM, Zhao X, et al. Secondary preventive medication persistence and adherence 1 year after stroke. *Neurology.* 2011; 77:1182–1190. [PubMed: 21900638]
21. Song X, Sander SD, Varker H, Amin A. Patterns and predictors of use of warfarin and other common long-term medications in patients with atrial fibrillation. *Am J Cardiovasc Drugs.* 2012; 12:245–253. [PubMed: 22775446]
22. Wieloch M, Sjölander A, Frykman V, Rosenqvist M, Eriksson N, Svensson PJ. Anticoagulation control in Sweden: reports of time in therapeutic range, major bleeding, and thrombo-embolic complications from the national quality registry Auricula. *Eur Heart J.* 2011; 32:2282–2289. [PubMed: 21616951]
23. Hylek EM, Heiman H, Skates SJ, Sheehan MA, Singer DE. Acetaminophen and other risk factors for excessive warfarin anticoagulation. *JAMA.* 1998; 279:657–662. [PubMed: 9496982]
24. Lip GY, Frison L, Halperin JL, Lane DA. Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: the HAS-BLED (Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile INR, Elderly, Drug). *J Am Coll Cardiol.* 2011; 57:173–180. [PubMed: 21111555]
25. Shireman TI, Mahnken JD, Howard PA, Kresowik TF, Hou Q, Ellerbeck EF. Development of a contemporary bleeding risk model for elderly warfarin recipients. *Chest.* 2006; 130:1390–1396. [PubMed: 17099015]
26. Beyth RJ, Quinn LM, Landefeld CS. Prospective evaluation of an index for predicting the risk of major bleeding in outpatients treated with warfarin. *Am J Med.* 1998; 105:91–99. [PubMed: 9727814]
27. Gage BF, Yan Y, Milligan PE, et al. Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF). *Am Hear J.* 2006; 151:713–719.
28. Kimmel SE, French B, Kasner SE, et al. A Pharmacogenetic versus a Clinical Algorithm for Warfarin Dosing. *N Engl J Med.* 2013; 369:2283–2293. [PubMed: 24251361]
29. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics.* 2000; 56:337–344. [PubMed: 10877287]

30. Steyerberg EW, Eijkemans MJC, Harrell FEJ, Habbema JDF. Prognostic modelling with logistic regression analysis : a comparison of selection and estimation methods in small data sets. *Stat Med.* 2000; 19:1059–1079. [PubMed: 10790680]
31. Liu M, Kapadia AS, Etzel CJ. Evaluating a New Risk Marker’s Predictive Contribution in Survival Models. *J Stat Theory Pract.* 2010; 4:845–855. [PubMed: 22984361]
32. Baker SG. Putting risk prediction in perspective: relative utility curves. *J Natl Cancer Inst.* 2009; 101:1538–1542. [PubMed: 19843888]
33. Baker SG, Cook NR, Vickers A, Kramer BS. Using relative utility curves to evaluate risk prediction. *J R Stat Soc Ser A.* 2009; 172:729–748.
34. R Development Core Team. R: A Language and Environment for Statistical Computing. 2014 [Accessed March 17, 2016] Available at: <http://www.r-project.org/>.
35. Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994; 81:515–526.
36. Bleeker SE, Moll HA, Steyerberg EW, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol.* 2003; 56:826–832. [PubMed: 14505766]
37. Kerr KF, McClelland RL, Brown ER, Lumley T. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. *Am J Epidemiol.* 2011; 174:364–374. [PubMed: 21673124]
38. Pepe MS, Feng Z, Gu JW. Comments on “Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond” by M. J. Pencina et al., *Statistics in Medicine* (DOI: 10.1002/sim.2929). *Stat Med.* 2008; 27:173–181. [PubMed: 17671958]
39. Finkelman BS, French B, Kimmel SE. The prediction accuracy of dynamic mixed-effects models in clustered data. *BioData Min.* 2016; 9:5. [PubMed: 26819631]
40. Hickey GL, Grant SW, Caiado C, et al. Dynamic prediction modeling approaches for cardiac surgery. *Circ Cardiovasc Qual Outcomes.* 2013; 6:649–658. [PubMed: 24150044]
41. McCormick TH, Raftery AE, Madigan D, Burd RS. Dynamic logistic regression and dynamic model averaging for binary classification. *Biometrics.* 2012; 68:23–30. [PubMed: 21838812]
42. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med.* 2015; 162:W1–W73. [PubMed: 25560730]
43. Li RC, Finkelman BS, Chen J, et al. Dietary vitamin K intake and anticoagulation control during the initiation phase of warfarin therapy: a prospective cohort study. *Thromb Haemost.* 2013; 110:195–196. [PubMed: 23572189]
44. Rombouts EK, Rosendaal FR, van der Meer FJM. Influence of dietary vitamin K intake on subtherapeutic oral anticoagulant therapy. *Br J Haematol.* 2010; 149:598–605. [PubMed: 20151978]

Key Points

- Clinical, social, and genetic factors are known to predict warfarin dose; however, it is unknown whether such factors predict prolonged dose titration for patients initiating warfarin therapy.
- This study aimed to derive a prediction model for prolonged dose titration in a prospective cohort of patients starting warfarin therapy and then validate the model in an external cohort.
- The prediction model did not perform well upon external validation. Accurate prediction of prolonged dose titration in warfarin patients remains unlikely with traditional clinical, social, and genetic factors.
- Future models for predicting prolonged dose titration will likely need to use methods that can accommodate substantial variation across clinics and over time.

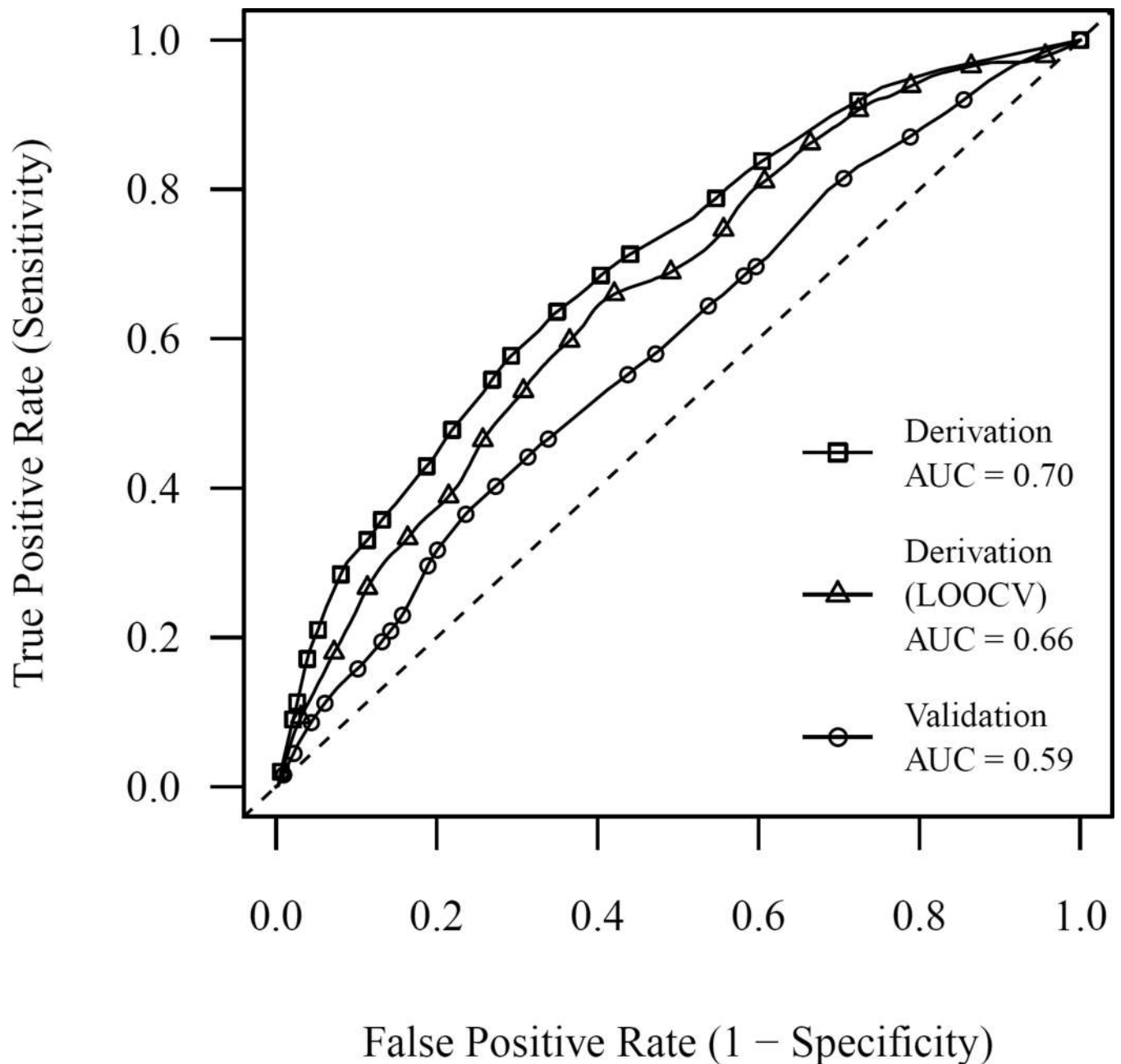


Figure 1.

ROC curve for the prediction model as tested in the derivation and validation cohorts. Curves are shown for the model tested in the derivation cohort, estimated using standard methods and leave-one-out cross-validation (LOOCV), as well as in the validation cohort, estimated using standard methods.

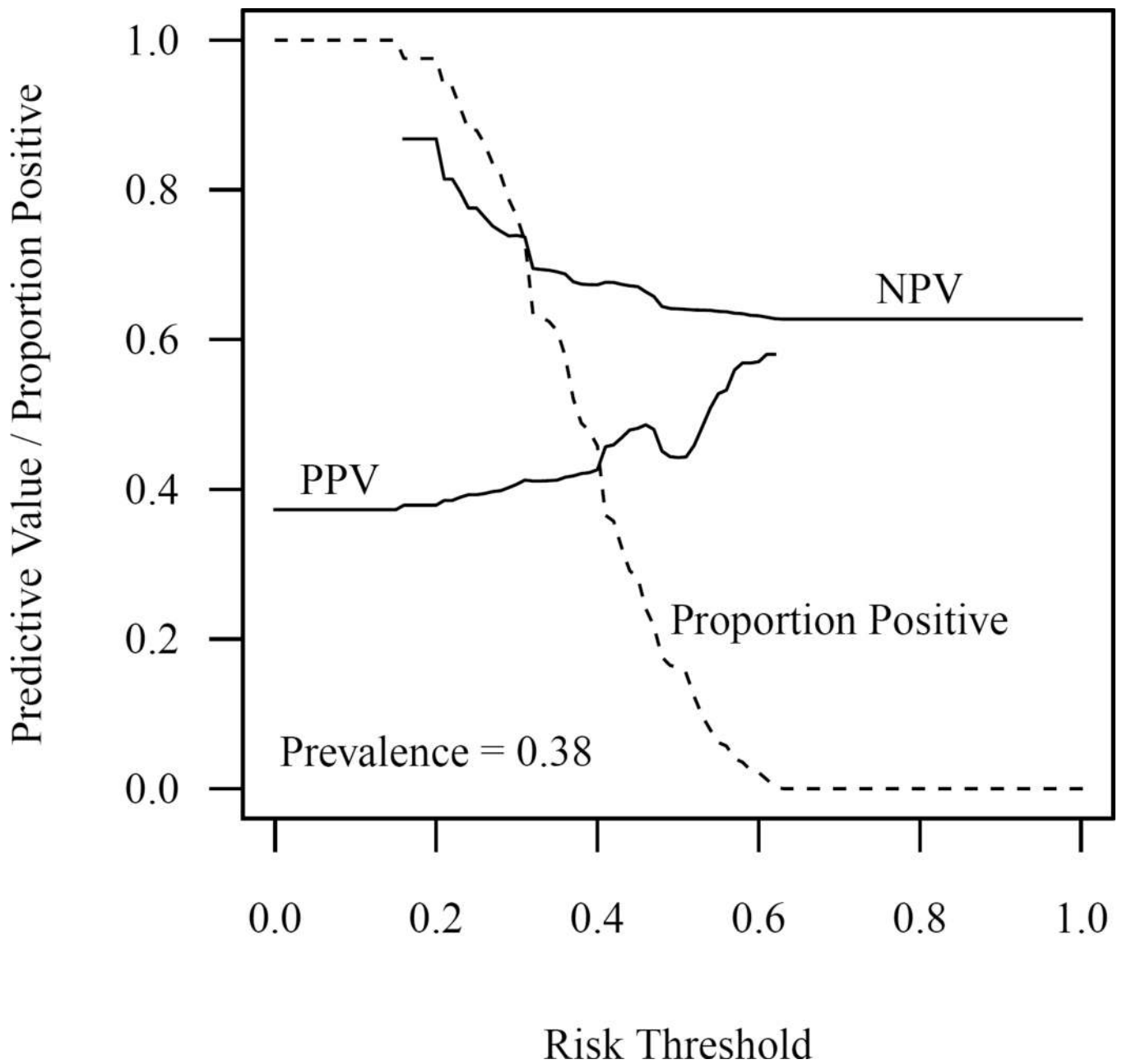


Figure 2.

Positive predictive value, negative predictive value, and proportion of patients classified as positive across the range of values for the risk threshold. Individuals with a predicted probability of prolonged dose titration are classified as positive. The absence of a curve in a given region indicates that the measure is undefined in that region; for instance, positive predictive value is undefined when no patients are classified as positive.

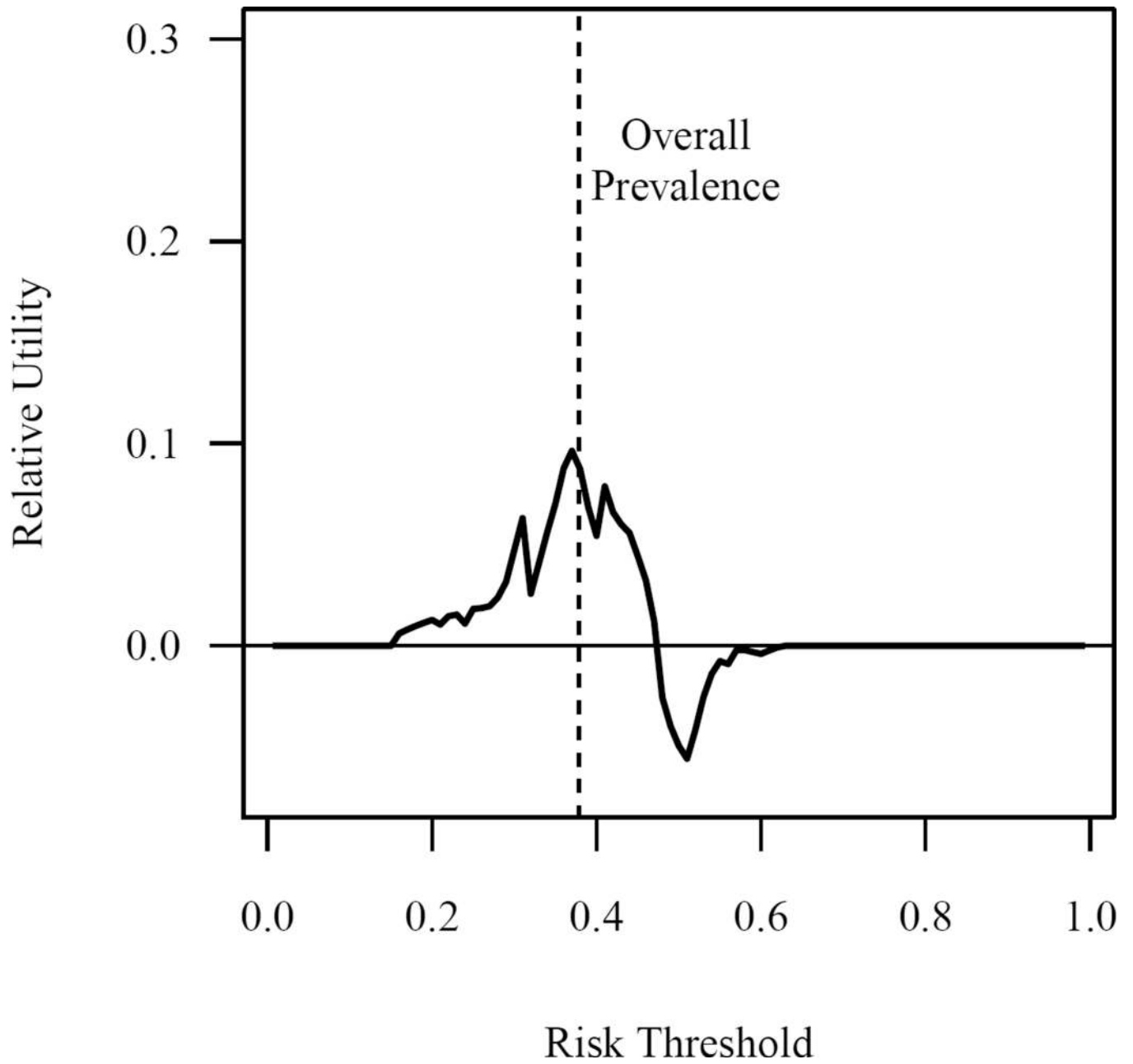


Figure 3.
Relative utility of the prediction model across the full range of risk thresholds.

Table 1

Characteristics of the derivation and validation cohorts

Variable	Derivation* (N = 390)	Validation* (N = 663)	P value†
Age			
< 45	65 (17)	135 (20)	< 0.001
45 – 55	74 (19)	131 (20)	
55 – 65	103 (26)	219 (33)	
65 – 75	83 (21)	116 (18)	
75+	65 (17)	60 (9)	
Female gender	119 (31)	250 (38)	0.02
African American race	174 (45)	466 (71)	< 0.001
Employment status			
Working	128 (33)	167 (25)	< 0.001
Unemployed	34 (9)	49 (7)	
Retired	143 (37)	192 (29)	
Disabled	81 (21)	251 (38)	
Annual income			
< \$15,000	109 (33)	228 (41)	< 0.001
\$15,000 – \$20,000	99 (30)	45 (8)	
> \$20,000	122 (37)	282 (51)	
Insurance status			
Private	215 (56)	276 (42)	< 0.001
VA/Medicare/Other	124 (32)	272 (41)	
Medicaid/None	45 (12)	110 (17)	
Number MD visits in previous year			
< 4	101 (26)	122 (19)	< 0.001
4 – 12	187 (48)	295 (45)	
> 12	98 (25)	242 (37)	
Smoking status			
Never	141 (36)	275 (42)	< 0.001
Past	185 (47)	235 (36)	
Current	64 (16)	148 (22)	
Body Mass Index			
< 25	122 (32)	185 (28)	0.11
25 – 30	125 (32)	189 (29)	
> 30	140 (36)	280 (43)	
Warfarin indication			
AFib/AFflutter	188 (48)	214 (32)	< 0.001
DVT/PE	116 (30)	343 (52)	
Other	86 (22)	105 (16)	
Previously used warfarin	96 (25)	209 (32)	0.02

Variable	Derivation* (N = 390)	Validation* (N = 663)	P value [†]
History of hypertension	192 (49)	461 (70)	< 0.001
History of diabetes	107 (27)	190 (29)	0.71
History of peptic ulcer disease	36 (9)	98 (15)	0.01
History of congestive heart failure	78 (20)	141 (21)	0.65

AFib indicates atrial fibrillation; AFlutter, atrial flutter; DVT, deep vein thrombosis; VA, Veterans Affairs.

* All values are reported as N (%)

[†] P values are based on the chi-square test.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Prediction model coefficients

Predictor variable	Shrunk coefficient*
Warfarin indication	
AFib/Aflutter	Ref
DVT/PE	-0.47
Other	-0.33
Insurance status	
Private insurance	Ref
VA/Medicare	-0.14
Medicaid/None	-0.42
Number MD visits in previous year	
<4	-0.29
4-12	Ref
>12	-0.23
Current smoker	-0.17
History of heart failure	-0.21

* To improve expected model calibration, coefficients were shrunk using a linear shrinkage factor, equal to 0.82, which was estimated from 1,000 bootstrap replications. Negative coefficients indicate a higher probability of prolonged dose titration.

Table 3

Model characteristics at various risk thresholds

Risk threshold*	C_{FP}/C_{FN}^{\dagger}	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Proportion Predicted Positive
10%	0.11	1.00	0.00	0.37	—	1.00
20%	0.25	0.99	0.03	0.38	0.87	0.98
30%	0.43	0.83	0.28	0.41	0.74	0.76
40%	0.67	0.52	0.58	0.43	0.67	0.46
50%	1	0.19	0.86	0.44	0.64	0.16
60%	1.5	0.03	0.99	0.57	0.63	0.02
70%	2.33	0.00	1.00	—	0.63	0.00

* The risk threshold refers to the cut-off probability for the outcome, where one classifies individuals as positive when predicted to be above the cut-off or negative when predicted to be below the cut-off. In this case, being “positive” refers to having a high probability of prolonged dose titration on warfarin, potentially leading a physician to choose an alternative therapy.

† C_{FP}/C_{FN} refers to the ratio of the costs of false positive and false negative mistakes that are implied by the risk threshold, according to decision theory.