METHODOLOGICAL ISSUES IN PREDICTION MODELS AND DATA ANALYSES USING OBSERVATIONAL AND CLINICAL TRIAL DATA

By

GUOWEI LI, MBBS, MSc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2016)

Department of Clinical Epidemiology and Biostatistics, Health Research Methodology, McMaster University

Hamilton, Ontario, Canada

TITLE: Methodological Issues in Prediction Models and Data Analyses Using Observational and Clinical Trial Data

AUTHOR: Guowei Li, MBBS (Peking University, 2010), MSc (Peking University, 2012)

SUPERVISOR: Dr. Lehana Thabane

NUMBER OF PAGES: xii, 106

ABSTRACT

Background and objectives

Prediction models are useful tools in clinical practise by providing predictive estimates of outcome probabilities to aid in decision making. As biomedical research advances, concerns have been raised regarding combined effectiveness (benefit) and safety (harm) outcomes in a prediction model, while typically different prediction models only focus on predictions of separate outcomes. A second issue is that, evidence also reveals poor predictive accuracy in different populations and settings for some prediction models, requiring model calibration or redevelopment. A third issue in data analyses is whether the treatment effect estimates could be influenced by competing risk bias. If other events preclude the outcomes of interest, these events would compete with the outcomes and thus fundamentally change the probability of the outcomes of interest. Failure to recognize the existence of competing risk or to account for it may result in misleading conclusions in health research. Therefore in this thesis, we explored three methodological issues in prediction models and data analyses by: (1) developing and externally validating a prediction model for patients' individual combined benefit and harm outcomes (stroke with no major bleeding, major bleeding with no stroke, neither event, or both stroke and major bleeding) with and without warfarin therapy for atrial fibrillation; (2) constructing a prediction model for hospital mortality in medical-surgical critically ill patients; and (3) performing a competing risk analysis to assess the efficacy of the low molecular weight heparin dalteparin versus unfractionated heparin in venous thromboembolism in medical-surgical critically ill patients.

Methods

Project 1: Using the Kaiser Permanente Colorado (KPCO) anticoagulation management cohort in the Denver-Boulder metropolitan area of Colorado in the United States to include patients with AF who were and were not prescribed warfarin therapy, we used a new approach to build a prediction model of individual combined benefit and harm outcomes related to warfarin therapy (stroke with no major bleeding, major bleeding with no stroke,

neither event, or both stroke and major bleeding) in patients with AF. We utilized a polytomous logistic regression (PLR) model to identify risk factors and then construct the new prediction model. Model performances and validation were evaluated systematically in the study.

Project 2: We used data from a multicenter randomized controlled trial named Prophylaxis for Thromboembolism in Critical Care Trial (PROTECT) to develop a new prediction model for hospital mortality in critically ill medical-surgical patients receiving heparin thromboprophylaxis. We first identified risk factors independent of APACHE (Acute Physiology and Chronic Health Evaluation) II score for hospital mortality, and then combined the identified risk factors and APACHE II score to build the new prediction model. Model performances were compared between the new prediction model and the APACHE II score.

Project 3: We re-analyzed the data from PROTECT to perform a sensitivity analysis based on a competing risk analysis to investigate the efficacy of dalteparin versus unfractionated heparin in preventing venous thromboembolism in medical-surgical critically ill patients, taking all-cause death as a competing risk for venous thromboembolism. Results from the competing risk analysis were compared with findings from the cause-specific analysis.

Results and Conclusions

Project 1: The PLR model could simultaneously predict risk of individual combined benefit and harm outcomes in patients with and without warfarin therapy for AF. The prediction model was a good fit, had acceptable discrimination and calibration, and was internally and externally validated. Should this approach be validated in other patient populations, it has potential advantages over existing risk stratification approaches.

Project 2: The new model combining other risk factors and APACHE II score was a good fit, well calibrated and internally validated. However, the discriminative ability of the prediction model was not satisfactory. Compared with the APACHE II score alone, the new prediction model increased data collection, was more complex but did not substantially improve

discriminative ability.

Project 3: The competing risk analysis yielded no significant effect of dalteparin compared with unfractionated heparin on proximal leg deep vein thromboses, but a lower risk of pulmonary embolism in critically ill medical-surgical patients. Findings from the competing risk analysis were similar to results from the cause-specific analysis.

PREFACE

This thesis is a ‘sandwich’ thesis, which combines three individual projects with four manuscripts that are either published or accepted by peer-reviewed journals. Results of the second and third projects have been published. For the first project, the protocol has been published while the main report is just accepted by a peer-reviewed journal. In this dissertation, the contributions of Guowei Li in all the papers included study conception, research question identification, study design, data analyses, interpretation of findings, manuscript writing, submitting the manuscripts, and responding to reviewers’ comments. My co-authors contributed to the acquisition of the datasets, assistance in computer programming, provision of clinical expertise, and critical revision of the manuscripts. The work of this thesis was conducted between Winter 2012 and Winter 2015.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Lehana Thabane. During my four-year Phd study, he has been providing prominent encouragement, guidance, support and caring. I am extremely thankful for all the advice, help and opportunities he provided especially when I just arrived in Canada and tired to adapt myself to the new life. I hope he will continue to be my life-long mentor and friend.

I am also very grateful to Dr. Anne Holbrook and Dr. Mitchell Levine for their endless and timely contributions to the projects presented in this thesis. The projects would never be finished without their invaluable and consistent suggestions, support and efforts.

Another big thank-you is given to Dr. Deborah Cook. She has provided tremendous support and encouragement in my research. Also many thanks are given to all my friends and colleagues from St Joseph's Healthcare Hamilton and McMaster University for their continuous help and support.

I am forever indebted to my beloved mother, Jieying Ou for her selfless love and support. I will forever miss and thank my lost father, Bisheng Li who could not share my achievements at McMaster University. I would like to thank my brothers Shijun Li and Jianwei Li who give me endless support. Last but not least, I would like to thank my wife Zhixiong Cheng and my little angel Anrui Li for their unconditional support, trust and love.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF FIGURES

CHAPTER 3

CHPATER 4

CHAPTER 5

Chapter 1

INTRODUCTION

Prediction models are useful tools in clinical practise by providing predictive estimates of outcome probabilities to aid in decision making.[1] Validated prediction models have been widely applied to inform individuals of their risk of developing illness and to guide patient-physician shared decisions on further interventions.[2, 3] Data used for the development of a prediction model should ideally be from a prospective cohort, or a randomized controlled trial.[2]

As biomedical research advances, an issue has been raised regarding the ability to combine effectiveness (benefit) and safety (harm) outcomes in a prediction model, while typically prediction models only focus on predictions of separate outcomes.[4] A second issue is that, evidence reveals poor predictive accuracy in different populations and settings for some prediction models, requiring model calibration or redevelopment.[5] A third issue in data analyses is whether the treatment effect estimates could be influenced by competing risk bias. If other events preclude the outcomes of interest, these events would compete with the outcomes and thus fundamentally change the probability of the outcomes of interest.[6, 7] Failure to recognize the existence of competing risk or to account for it may result in misleading conclusions in health research.[8]

To address these issues, the objectives of this thesis are: (1) to develop and externally validate a prediction model for patients' individual combined benefit and harm outcomes (stroke with no major bleeding, major bleeding with no stroke, neither event, or both stroke and major bleeding) with and without warfarin therapy for atrial fibrillation; (2) to construct a prediction model for hospital mortality in medical-surgical critically ill patients; and (3) to perform a competing risk analysis to assess the efficacy of the low molecular weight heparin dalteparin versus unfractionated heparin in venous thromboembolism in medical-surgical critically ill patients.

Issue 1: Prediction of individual combined benefit and harm outcomes related to warfarin therapy

Atrial fibrillation (AF) is a major and independent risk factor for stroke and mortality. [9, 10] Despite the increased use of non-vitamin K antagonist oral anticoagulants, warfarin remains the mainstay for stroke prophylaxis in patients with AF.[10-12] Nevertheless, the use of warfarin is associated with an increased risk of major bleeding including intracranial hemorrhage (ICH), which may have resulted in the underuse by patients with AF who were qualified candidates for warfarin therapy.[13-17]

There have been validated clinical prediction models to assist physicians with predicted estimates of patients' risks of stroke and major bleeding. These include the **$CHADS_2$** (**C**ongestive heart failure, **H**ypertension, **A**ge > 75 years, **D**iabetes, Previous **s**troke [2 points]) and the **CHA_2DS_2-VASc** scores (**C**ongestive heart failure; **H**ypertension; **A**ge ≥ 75 years [2 points]; **D**iabetes mellitus; **S**troke [2 points], **V**ascular disease, **A**ge 65-74 years, and **S**ex [female]) for risk of stroke,[10, 18-20] and the **HAS-BLED** score (**H**ypertension; **A**bnormal renal/liver function; **S**troke; **B**leeding history or predisposition; **L**abile international normalized ratio [INR], **E**lderly [>65 years]; **D**rugs/alcohol concomitantly) for risk of major bleeding.[10, 20, 21] Nevertheless, these risk-stratification tools cannot provide the individual combined benefit and harm probabilities needed by patients and physicians when warfarin therapy initiation is under consideration. There are several studies using the 'net benefit' approach to take into account stroke and major bleeding simultaneously; that is, Net Benefit = ($TE\ rate_{off\ warfarin}$ - $TE\ rate_{on\ warfarin}$) - Weight * ($ICH\ rate_{\ on\ warfarin}$ - $ICH\ rate_{\ off\ warfarin}$), where TE denotes ischemic stroke or systemic embolism, and ICH denotes intracranial hemorrhage.[22-24] However, this approach does not consider gastrointestinal bleeding, and the weighting factor reflecting the importance of ICH is identified arbitrarily and subjectively. In addition, some studies have tried to combine stroke and bleeding risk-stratification scores to estimate overall clinical outcome probabilities including stroke and major bleeding.[25, 26] Notwithstanding, the combined risk-stratification scores cannot improve prediction of stroke

and major bleeding beyond the individual stroke ($CHADS_2$, CHA_2DS_2-VASc) or bleeding (HAS-BLED) scores.[4]

Thus this study used a new approach to develop and validate a prediction model for patients' individual combined benefit and harm outcomes (stroke with no major bleeding, major bleeding with no stroke, neither event, or both stroke and major bleeding) with and without warfarin therapy for AF. In the model building, we also assessed the impact of the competing risk of death for the combined benefit and harm outcomes on the construction of the new prediction model. We used the Kaiser Permanente Colorado (KPCO) anticoagulation management cohort in the Denver-Boulder metropolitan area of Colorado in the United States to include patients with AF who were and were not prescribed warfarin therapy for analyses.[27] This new prediction model of individualized combined benefit and harm outcomes may aid in the patient-physician shared decision-making process for consideration of warfarin therapy initiation in real-world settings.

Issue 2: Risk factors for and prediction of mortality in critically ill medical-surgical patients receiving heparin thromboprophylaxis

Despite advances in medicine, mortality rates in critically ill patients remain substantially high.[28-30] The Acute Physiology and Chronic Health Evaluation (APACHE) prognostic scoring system has been used worldwide to evaluate the severity of illness and estimate the risk of hospital mortality for critically ill patients, based on the data collected in the first 24 hours of an intensive care unit (ICU) admission.[31-36] Though the APACHE has developed four generations of models,[37-40] the APACHE II score remains the most commonly used severity scoring tool in clinical practice and health research mainly due to its simplicity.[41] However, evidence shows that there is limited accuracy of APACHE II score for predicting the risk of mortality in different populations and countries.[42-46] For instance, one systematic review found that the median area under the receiver operating characteristic curve of APACHE II score for all-cause mortality in critically ill patients was only 0.77.[5] As well, emerging studies have reported additional risk factors for death which are independent of APACHE II scores,

including body mass index (BMI),[47, 48] sex,[49-51] use of vasopressors,[52] prothrombin index,[53, 54] and platelet count.[55, 56]

In order to improve the predictive accuracy of APACHE II score, the objective of this study was to identify risk factors independent of APACHE II score, and develop and validate a new prediction model for hospital mortality. Model performances were compared between the new prediction model and the APACHE II score. Data were used from a multicenter randomized controlled trial named Prophylaxis for Thromboembolism in Critical Care Trial (PROTECT) in critically ill medical-surgical patients receiving heparin thromboprophylaxis.[57] The new prediction model may be helpful for physicians in patient assessments and management considerations in critically ill patients.

Issue 3: Competing risk analysis for evaluation of dalteparin versus unfractionated heparin for venous thromboembolism in medical-surgical critically ill patients

In biomedical research, it is not uncommon for participants to develop a competing risk event which prevents observing the outcome of interest.[6, 7] The competing risk event would fundamentally alter the probability of the outcome of interest, thus leading to potential bias in treatment effect estimates in clinical trials or relationship estimates in epidemiological research.[8]

Critically ill patients admitted to intensive care units (ICUs) are at high risk of venous thromboembolism (VTE).[58-60] The multicenter international randomized controlled trial, PROTECT (Prophylaxis for Thromboembolism in Critical Care Trial), evaluated the efficacy of dalteparin (a low-molecular-weight heparin) versus unfractionated heparin in preventing proximal leg deep vein thromboses (PLDVT) and other VTEs in critically ill patients.[57] Using standard survival analysis (also known as cause-specific analysis), this trial found no significant effect of dalteparin versus unfractionated heparin on PLDVT, but a significantly superior treatment effect of dalteparin on pulmonary embolism.[57] Nevertheless, the all-cause mortality (23%) was significantly higher than the rate of PLDVT (6%) and pulmonary

embolism (2%) in the participants. Death preceding a VTE can preclude the occurrence of subsequent PLDVT and pulmonary embolism; therefore death is a competing risk event which may potentially bias the treatment effect estimates. It has been reported that the cause-specific analysis would yield biased findings because it fails to take into account competing risks.[8, 61, 62] Another concern is that the cause-specific analysis could be inappropriate for competing risk analysis since the assumptions of non-informative censoring and independence of time distributions between VTE and death may have been violated.[7, 63] For example, it is not appropriate to use the Kaplan-Meier method to estimate survival curves for VTE due to the competing risk of death. By contrast, the cumulative incidence function (CIF, also known as the subdistribution) which is derived from the cause-specific hazard function and does not require the independence assumption, should be used to estimate the marginal probability of VTE in the presence of competing risk.[64]

We therefore re-analyzed the data from PROTECT to perform a sensitivity analysis based on a competing risk analysis to investigate the efficacy of dalteparin versus unfractionated heparin in preventing VTE in medical-surgical critically ill patients, taking all-cause death as a competing risk for VTE. We also compared results from the competing risk analysis with findings from the cause-specific analysis.

Outline of the thesis

This thesis is a 'sandwich' of three projects corresponding with the three issues described above. The papers are separated into four different chapters beginning with Chapter 2.

Chapter 2 is the proposal for the prediction model of individual combined benefit and harm outcomes related to warfarin therapy, proposing all the details on the methodology used. In **Chapter 3**, we used a new approach to develop and validate a prediction model for patients' individual combined benefit and harm outcomes (stroke with no major bleeding, major bleeding with no stroke, neither event, or both stroke and major bleeding) with and without warfarin therapy for AF. We firstly identified risk factors and built a new prediction model for

prediction of the combined benefit and harm outcomes in patients with and without warfarin, based on polytomous logistic regression analysis. Subsequently, the model performances, internal and external validation were assessed systematically. The prediction model was also applied to predicting the combined benefit and harm outcomes with and without warfarin therapy for a typical patient newly diagnosed with AF.

In **Chapter 4**, we identified risk factors independent of APACHE II scores for hospital mortality in critically ill patients, and then developed a new prediction model that combined the APACHE II score with these additional risk factors. Model performances and validation for the new prediction model were evaluated. We also compared the model performances between the new prediction model and the APACHE II score.

In **Chapter 5**, a competing risk analysis was conducted to evaluate the efficacy of dalteparin versus unfractionated heparin in preventing VTE in medical-surgical critically ill patients, based on data from the PROTECT. The all-cause mortality was taken as the competing risk for VTE. Results from the competing risk analysis were compared with those from the cause-specific analysis.

Chapter 6 summarizes the main findings of Chapters 2 to 5. The implications and limitations of these three studies are also discussed in Chapter 6.

References:

1. Adams ST, Leveson SH. Clinical prediction rules. *BMJ (Clinical research ed.)*. 2012;344:d8312
2. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart (British Cardiac Society)*. 2012;98:683-690
3. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? *BMJ (Clinical research ed.)*. 2009;338:b375
4. Dzeshka MS, Lane DA, Lip GY. Stroke and bleeding risk in atrial fibrillation: Navigating the alphabet soup of risk-score acronyms (chads2 , cha2 ds2 -vasc, r2 chads2 , has-bled, atria, and more). *Clinical cardiology*. 2014;37:634-644
5. Siontis G, Tzoulaki I, Ioannidis J. Predicting death: An empirical evaluation of predictive tools for mortality. *Archives of internal medicine*. 2011:archinternmed. 2011.2334 v2011
6. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. *Statistics in medicine*. 1999;18:695-706
7. Putter H, Fiocco M, Geskus R. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in medicine*. 2007;26:2389-2430
8. Pintilie M. An introduction to competing risks analysis. *Revista Española de Cardiología (English Edition)*. 2011;64:599-605
9. Hart RG, Halperin JL, Pearce LA, Anderson DC, Kronmal RA, McBride R, Nasco E, Sherman DG, Talbert RL, Marler JR. Lessons from the stroke prevention in atrial fibrillation trials. *Ann Intern Med*. 2003;138:831-838
10. Camm AJ, Kirchhof P, Lip GY, Schotten U, Savelieva I, Ernst S, Van Gelder IC, Al-Attar N, Hindricks G, Prendergast B, Heidbuchel H, Alfieri O, Angelini A, Atar D, Colonna P, De Caterina R, De Sutter J, Goette A, Gorenek B, Heldal M, Hohloser SH, Kolh P, Le Heuzey JY, Ponikowski P, Rutten FH. Guidelines for the management of

atrial fibrillation: The task force for the management of atrial fibrillation of the european society of cardiology (esc). *Eur Heart J*. 2010;31:2369-2429

11. Rockson SG, Albers GW. Comparing the guidelines: Anticoagulation therapy to optimize stroke prevention in patients with atrial fibrillation. *J Am Coll Cardiol*. 2004;43:929-935
12. Wells G, Coyle D, Cameron C, Steiner S, Coyle K, Kelly S, Tang A, Healey J, Hsieh S, Berkom J. Safety, effectiveness, and cost-effectiveness of new oral anticoagulants compared with warfarin in preventing stroke and other cardiovascular events in patients with atrial fibrillation [internet]. Available from: Http://www.Cadth.Ca/media/pdf/noac_therapeutic_review_final_report.Pdf. *Ottawa: Canadian Agency for Drugs and Technologies in Health; 2012 Apr 9. 173 p. (CADTH Therapeutic Review)*.
13. Holbrook A, Labiris R, Goldsmith CH, Ota K, Harb S, Sebaldt RJ. Influence of decision aids on patient preferences for anticoagulant therapy: A randomized trial. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2007;176:1583-1587
14. Ansell J. Warfarin versus new agents: Interpreting the data. *Hematology Am Soc Hematol Educ Program*. 2010;2010:221-228
15. Arepally G, Bauer KA, Bhatt DL, Merli GJ, Naccarelli GV, Carter RD, Karcher RB, Berry CA, Keaton KL, Stowell SA. The use of antithrombotic therapies in the prevention and treatment of arterial and venous thrombosis: A survey of current knowledge and practice supporting the need for clinical education. *Crit Pathw Cardiol*. 2010;9:41-48
16. McCrory DC, Matchar DB, Samsa G, Sanders LL, Pritchett EL. Physician attitudes about anticoagulation for nonvalvular atrial fibrillation in the elderly. *Arch Intern Med*. 1995;155:277-281
17. Holbrook A, Pullenayegum E, Troyan S, Nikitovic M, Crowther M. Personalized benefit-harm information influences patient decisions regarding warfarin. *J Popul Ther Clin Pharmacol*. 2013;20:e406-e415
18. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk

stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: The euro heart survey on atrial fibrillation. *Chest*. 2010;137:263-272

19. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: Results from the national registry of atrial fibrillation. *JAMA*. 2001;285:2864-2870
20. Cairns JA, Connolly S, McMurtry S, Stephenson M, Talajic M. Canadian cardiovascular society atrial fibrillation guidelines 2010: Prevention of stroke and systemic thromboembolism in atrial fibrillation and flutter. *Can J Cardiol*. 2011;27:74-90
21. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (has-bled) to assess 1-year risk of major bleeding in patients with atrial fibrillation: The euro heart survey. *Chest*. 2010;138:1093-1100
22. Singer DE, Chang Y, Fang MC, Borowsky LH, Pomernacki NK, Udaltsova N, Go AS. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. *Ann Intern Med*. 2009;151:297-305
23. Friberg L, Rosenqvist M, Lip GY. Net clinical benefit of warfarin in patients with atrial fibrillation: A report from the swedish atrial fibrillation cohort study. *Circulation*. 2012;125:2298-2307
24. Olesen JB, Lip GY, Lindhardsen J, Lane DA, Ahlehoff O, Hansen ML, Raunso J, Tolstrup JS, Hansen PR, Gislason GH, Torp-Pedersen C. Risks of thromboembolism and bleeding with thromboprophylaxis in patients with atrial fibrillation: A net clinical benefit analysis using a 'real world' nationwide cohort study. *Thromb Haemost*. 2011;106:739-749
25. Lip GY, Lane DA, Buller H, Apostolakis S. Development of a novel composite stroke and bleeding risk score in patients with atrial fibrillation: The amadeus study. *Chest*. 2013;144:1839-1847
26. Banerjee A, Fauchier L, Bernard-Brunet A, Clementy N, Lip GY. Composite risk scores and composite endpoints in the risk prediction of outcomes in anticoagulated patients with atrial fibrillation. The loire valley atrial fibrillation project. *Thromb*

Haemost. 2014;111:549-556

27. Witt DM, Sadler MA, Shanahan RL, Mazzoli G, Tillman DJ. Effect of a centralized clinical pharmacy anticoagulation service on the outcomes of anticoagulation therapy. *Chest*. 2005;127:1515-1522
28. Checkley W, Martin GS, Brown SM, Chang SY, Dabbagh O, Fremont RD, Girard TD, Rice TW, Howell MD, Johnson SB. Structure, process, and annual icu mortality across 69 centers: United states critical illness and injury trials group critical illness outcomes study. *Critical care medicine*. 2013
29. Wu AW, Pronovost P, Morlock L. Icu incident reporting systems. *Journal of critical care*. 2002;17:86-94
30. Parker J. California intensive care outcomes project (calico). *Available at: http://www.oshpd.ca.gov/HID/Products/PatDischargeData/ICUDataCALICO/. Accessed February 20, 2015*. 2007
31. Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, de Jonge E, Bosman RJ, Peelen L, de Keizer NF. External validation of acute physiology and chronic health evaluation iv in dutch intensive care units and comparison with acute physiology and chronic health evaluation ii and simplified acute physiology score ii. *Journal of critical care*. 2011;26:105 e111-108
32. Zimmerman JE, Kramer AA. Outcome prediction in critical care: The acute physiology and chronic health evaluation models. *Current opinion in critical care*. 2008;14:491-497
33. Beck DH, Smith GB, Pappachan JV, Millar B. External validation of the saps ii, apache ii and apache iii prognostic models in south england: A multicentre study. *Intensive care medicine*. 2003;29:249-256
34. Wong DT, Knaus WA. Predicting outcome in critical care: The current status of the apache prognostic scoring system. *Canadian journal of anaesthesia = Journal canadien d'anesthesie*. 1991;38:374-383
35. OH TE, Hutchinson R, Short S, BUCKLEY T, LIN E, LEUNG D. Verification of the acute physiology and chronic health evaluation scoring system in a hong kong intensive care unit. *Critical care medicine*. 1993;21:698-705

36. Capuzzo M, Valpondi V, Sgarbi A, Bortolazzi S, Pavoni V, Gilli G, Candini G, Gritti G, Alvisi R. Validation of severity scoring systems saps ii and apache ii in a single-center population. *Intensive care medicine*. 2000;26:1779-1785

37. KNAUS WA, DRAPER EA, WAGNER DP, ZIMMERMAN JE, BIRNBAUM ML, CULLEN DJ, KOHLES MK, SHIN B, SNYDER JV. Evaluating outcome from intensive care: A preliminary multihospital comparison. *Critical care medicine*. 1982;10:491-496

38. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. Apache ii: A severity of disease classification system. *Critical care medicine*. 1985;13:818-829

39. Knaus WA, Wagner D, Draper Eea, Zimmerman J, Bergner M, Bastos PG, Sirio C, Murphy D, Lotring T, Damiano A. The apache iii prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *CHEST Journal*. 1991;100:1619-1636

40. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute physiology and chronic health evaluation (apache) iv: Hospital mortality assessment for today's critically ill patients*. *Critical care medicine*. 2006;34:1297-1310

41. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. *Continuing Education in Anaesthesia, Critical Care & Pain*. 2008;8:181-185

42. Arabi Y, Al Shirawi N, Memish Z, Venkatesh S, Al-Shimemeri A. Assessment of six mortality prediction models in patients admitted with severe sepsis and septic shock to the intensive care unit: A prospective cohort study. *Critical care*. 2003;7:R116

43. Keegan MT, Gali B, Findlay JY, Heimbach JK, Plevak DJ, Afessa B. Apache iii outcome prediction in patients admitted to the intensive care unit after liver transplantation: A retrospective cohort study. *BMC surgery*. 2009;9:11

44. Namendys-Silva SA, Baltazar-Torres JA, Rivero-Sigarroa E, Fonseca-Lazcano JA, Montiel-Lopez L, Dominguez-Cherit G. Prognostic factors in patients with systemic lupus erythematosus admitted to the intensive care unit. *Lupus*. 2009;18:1252-1258

45. Papachristou GI, Muddana V, Yadav D, O'Connell M, Sanders MK, Slivka A, Whitcomb DC. Comparison of bisap, ranson's, apache-ii, and ctsi scores in predicting organ failure, complications, and mortality in acute pancreatitis. *The American*

journal of gastroenterology. 2010;105:435-441; quiz 442

46. Inal MT, Memis D, Kargi M, Sut N. Prognostic value of indocyanine green elimination assessed with limon in septic patients. *Journal of critical care.* 2009;24:329-334
47. O'Brien Jr JM, Phillips GS, Ali NA, Lucarelli M, Marsh CB, Lemeshow S. Body mass index is independently associated with hospital mortality in mechanically ventilated adults with acute lung injury. *Critical care medicine*. 2006;34:738
48. Pickkers P, de Keizer N, Dusseljee J, Weerheijm D, van der Hoeven JG, Peek N. Body mass index is associated with hospital mortality in critically ill patients: An observational cohort study. *Crit Care Med*. 2013;41:1878-1883
49. Mahmood K, Eldeirawi K, Wahidi MM. Association of gender with outcomes in critically ill patients. *Crit Care*. 2012;16:R92
50. Combes A, Luyt CE, Trouillet JL, Nieszkowska A, Chastre J. Gender impact on the outcomes of critically ill patients with nosocomial infections. *Critical care medicine*. 2009;37:2506-2511
51. Nachtigall I, Tafelski S, Rothbart A, Kaufner L, Schmidt M, Tamarkin A, Kartachov M, Zebedies D, Trefzer T, Wernecke KD, Spies C. Gender-related outcome difference is related to course of sepsis on mixed icus: A prospective, observational clinical study. *Crit Care*. 2011;15:R151
52. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR. Saps 3--from evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at icu admission. *Intensive care medicine*. 2005;31:1345-1355
53. Walsh TS, Stanworth SJ, Prescott RJ, Lee RJ, Watson DM, Wyncoll D. Prevalence, management, and outcomes of critically ill patients with prothrombin time prolongation in united kingdom intensive care units. *Critical care medicine*. 2010;38:1939-1946
54. Angstwurm MW, Dempfle CE, Spannagl M. New disseminated intravascular coagulation score: A useful tool to predict mortality in comparison with acute physiology and chronic health evaluation ii and logistic organ dysfunction scores.

Critical care medicine. 2006;34:314-320; quiz 328

55. Williamson DR, Albert M, Heels-Ansdell D, Arnold DM, Lauzier F, Zarychanski R, Crowther M, Warkentin TE, Dodek P, Cade J, Lesur O, Lim W, Fowler R, Lamontagne F, Langevin S, Freitag A, Muscedere J, Friedrich JO, Geerts W, Burry L, Alhashemi J, Cook D. Thrombocytopenia in critically ill patients receiving thromboprophylaxis: Frequency, risk factors, and outcomes. *Chest*. 2013;144:1207-1215
56. Strauss R, Wehler M, Mehler K, Kreutzer D, Koebnick C, Hahn EG. Thrombocytopenia in patients in the medical intensive care unit: Bleeding prevalence, transfusion requirements, and outcome. *Critical care medicine*. 2002;30:1765-1771
57. Cook D, Meade M, Guyatt G, Walter S, Heels-Ansdell D, Warkentin TE, Zytaruk N, Crowther M, Geerts W, Cooper DJ. Dalteparin versus unfractionated heparin in critically ill patients. *The New England journal of medicine*. 2011;364:1305-1314
58. Geerts WH, Bergqvist D, Pineo GF, Heit JA, Samama CM, Lassen MR, Colwell CW. Prevention of venous thromboembolism: American college of chest physicians evidence-based clinical practice guidelines (8th edition). *Chest*. 2008;133:381S-453S
59. Cook D, Crowther M, Meade M, Rabbat C, Griffith L, Schiff D, Geerts W, Guyatt G. Deep venous thrombosis in medical-surgical critically ill patients: Prevalence, incidence, and risk factors. *Critical care medicine*. 2005;33:1565-1571
60. Geerts W, Selby R. Prevention of venous thromboembolism in the icu. *CHEST Journal*. 2003;124:357S-363S
61. Forsblom C, Harjutsalo V, Thorn LM, Wadén J, Tolonen N, Saraheimo M, Gordin D, Moran JL, Thomas MC, Groop P-H. Competing-risk analysis of esrd and death among patients with type 1 diabetes and macroalbuminuria. *Journal of the American Society of Nephrology*. 2011;22:537-544
62. Berry SD, Ngo L, Samelson EJ, Kiel DP. Competing risk of death: An important consideration in studies of older adults. *Journal of the American Geriatrics Society*. 2010;58:783-787
63. Resche-Rigon M, Azoulay E, Chevret S. Evaluating mortality in intensive care units: Contribution of competing risks analyses. *Critical Care*. 2005;10:R5

64. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999;94:496-509

Chapter 2

BMJ Open

Prediction of individual combined benefit and harm for patients with atrial fibrillation considering warfarin therapy: a study protocol

Guowei Li,[1] Anne Holbrook,[1,2,3] Thomas Delate,[4,5] Daniel M Witt,[6] Mitchell AH Levine,[1,2,3] Lehana Thabane[1,2]

To cite: Li G, Holbrook A, Delate T, *et al*. Prediction of individual combined benefit and harm for patients with atrial fibrillation considering warfarin therapy: a study protocol. *BMJ Open* 2015;**5**: e009518. doi:10.1136/bmjopen-2015-009518

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2015-009518).

Received 24 July 2015
Revised 25 September 2015
Accepted 14 October 2015

For numbered affiliations see end of article.

Correspondence to
Professor Lehana Thabane; thabanl@mcmaster.ca

ABSTRACT

Introduction: Clinical prediction rules have been validated and widely used in patients with atrial fibrillation (AF) to predict stroke and major bleeding. However, these prediction rules were not developed in the same population, and do not provide the key information that patients and prescribers need at the time anticoagulants are being considered—what is the individual patient-specific risk of both benefit (decreased stroke) and harm (increased major bleeding). In this study, our primary objective is to develop and validate a prediction model for patients' individual combined benefit and harm outcomes (stroke, major bleeding and neither event) with and without warfarin therapy. Our secondary outcome is all-cause mortality.

Methods and analysis: We will use data from the Kaiser Permanente Colorado (KPCO) anticoagulation management databases and electronic medical records. Patients with a primary or secondary diagnosis during an ambulatory KPCO medical office visit, emergency department visit, or inpatient stay between 1 January 2005 and 31 December 2012 with no AF diagnosis in the previous 180 days will be included. Patients' demographic characteristics, laboratory data, comorbidities, warfarin medication data and concurrent use of medication will be used to construct the prediction model. For primary outcomes (stroke with no major bleeding, and major bleeding with no stroke), we will perform polytomous logistic regression to develop a prediction model for patients' individual combined benefit and harm outcomes, taking neither event group as the reference group. As regards death, we will use Cox proportional hazards regression analysis to build a prediction model for all-cause mortality.

Ethics and dissemination: This study has been approved by the KPCO Institutional Review Board and the Hamilton Integrated Research Ethics Board. Results from this study will be published in a peer-reviewed journal electronically and in print. The prediction models may aid in patient-physician shared decision-making when they are considering warfarin therapy.

Strengths and limitations of this study

- The prediction model can provide comprehensive information on the individual combined benefit and harm with and without warfarin for each patient with atrial fibrillation.
- The prediction model may aid in patient-physician shared decision-making when they are considering warfarin therapy.
- Rigorous statistical analyses are performed for model construction and assessment.
- Potential limitation includes the data accuracy of the administrative databases used in this study.

BACKGROUND

Atrial fibrillation (AF) is the most common sustained cardiac dysrhythmia. The presence of AF is a strong and independent risk factor for stroke[1] with an approximate fivefold excess risk,[2] and for mortality with a doubled death rate.[3] Antithrombotic therapies such as oral warfarin are now the mainstay for stroke prevention and recommended in guidelines for patients with AF.[1 3–5]

Warfarin is impressively efficacious in preventing stroke and death. A recent meta-analysis of randomised controlled trials concluded that warfarin reduced the risk of stroke and mortality by 64% and 26%, respectively, compared with placebo or no treatment.[6] Several new oral anticoagulants including dabigatran, rivaroxaban and apixaban are now available. Their use is increasing, primarily because they do not require routine anticoagulation intensity monitoring such as the international normalised ratio (INR).[7] At present, while the evidence, especially real-world clinical practice evidence, is evolving for these new drugs, warfarin remains the dominant oral anticoagulant.[8] For all anticoagulants, the most important

and worrisome adverse event is major bleeding especially intracranial haemorrhage (ICH).[5 9 10] Fear of bleeding, risk proclivity regarding stroke, and antipathy about taking an additional medication and undergoing blood tests has led to underuse by patients with AF who are qualified candidates for warfarin therapy.[11–15]

Clinical prediction rules have been validated in patients with AF to predict stroke and bleeding (JA Pereira. Methods to predict individualized combined benefit/harm patient profiles for warfarin. Graduate Department of Pharmaceutical Sciences [unpublished doctoral dissertation] Toronto, Canada: University of Toronto, 2008). The $CHADS_2$ (*C*ongestive heart failure, *H*ypertension, *A*ge >75 years, *D*iabetes, Previous *S*troke (2 points)) and the CHA_2DS_2-VASc scores are recommended (*C*ongestive heart failure; *H*ypertension; *A*ge ≥75 years (2 points); *D*iabetes mellitus; *S*troke (2 points); *V*ascular disease, *A*ge 65–74 years and *S*ex (female)) to predict the risk of stroke.[3 5 16 17] For prediction of major bleeding with warfarin, the HAS-BLED score (*H*ypertension; *A*bnormal renal/liver function; *S*troke; *B*leeding history or predisposition; *L*abile INR; *E*lderly (>65 years); *D*rugs/alcohol concomitantly) has been validated.[3 5 18–21]

Despite their usefulness as guides, the classification schemes above were not developed in the same population, and do not provide the key information that patients and prescribers need at the time anticoagulants are being considered—what is the individual patient-specific risk of both benefit (decreased stroke) and harm (increased bleeding). A prediction rule to assess the probabilities of both stroke and major bleeding simultaneously in the same population is required. Some studies have focused on combined benefit and harm profiles of warfarin versus no warfarin for individual patients with newly diagnosed AF. For example, several studies have used a 'net benefit' approach for warfarin which took into account the main benefit (reduced risk of stroke) and the main harm (increased risk of bleeding) in the same population.[22–24] However, the weighting factor reflecting relative impact for calculation (ie, net benefit=$(\text{TE rate}_{\text{off warfarin}} - \text{TE rate}_{\text{on warfarin}}) - \text{weight} \times (\text{ICH rate}_{\text{on warfarin}} - \text{ICH rate}_{\text{off warfarin}})$, where TE denotes ischaemic stroke or systemic embolism) was chosen arbitrarily.[22–24] The chosen weight of 1.5 for ICH assumes that patients with AF would weigh ICH as 50% worse than TE. These studies failed to give careful consideration to actual data on patient values and preferences,[25–27] and did not consider the much more common gastrointestinal (GI) bleeding.

Considerations of rates of benefit and harm in patients with AF must also consider mortality, since these are typically older patients and mortality is high. For instance, in an American community-based cohort, the death rate in a population with AF was over 60% during a mean follow-up of 5.3 years.[28] Therefore, death is essentially a competing risk of stroke and major bleeding. However, the existing risk stratification schemes do not deal with the death as a competing risk.[16–18 29] Given that the three non-lethal outcomes associated with AF ('no major bleeding and no stroke', 'stroke' and 'major bleeding') are not mutually exclusive, and patients deserve to have this critical information on their individual risk of benefit and harm at the time they are considering whether to take warfarin or not, it is imperative to derive valid predictions for and calculate the probabilities of individualised combined benefit and harm outcomes of warfarin. Furthermore, methods of considering mortality as a competing risk with stroke and major bleeding are needed.

In this study, we will complete the development and external validation of a prediction model for each patient's individual combined benefit and harm outcomes (stroke with no major bleeding, major bleeding with no stroke and neither event) with and without warfarin therapy, based on the Kaiser Permanente Colorado (KPCO) anticoagulation management databases and electronic medical records. Our secondary objective is to devise a model to predict all-cause mortality.

METHODS

Patients and settings

KPCO is a geographic section of a large national non-profit group model Health Maintenance Organization. KPCO is an integrated healthcare delivery system providing medical care to approximately 550 000 patients in the Denver-Boulder metropolitan area of Colorado in the USA.[30] Patients who are members of KPCO are offered anticoagulation services by a centralised Clinical Pharmacy Anticoagulation Service (CPAS) (JA Pereira. Unpublished doctoral dissertation, 2008). CPAS clinical pharmacists initiate, adjust and refill warfarin and order relevant laboratory tests for patients, working in collaboration with the referring physicians and applying standardised dosing algorithms.[30] In this study, data will also be obtained for patients with AF in KPCO who are not taking warfarin or managed by CPAS.

New diagnoses of AF will be determined using International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes 427.31 and 427.32 from the KPCO Virtual Data Warehouse (VDW) Diagnosis Database. The codes can be recorded either as primary or secondary diagnosis during an ambulatory KPCO medical office visit, emergency department (ED) visit or inpatient stay between 1 January 2005 and 31 December 2012 with no AF diagnosis in the previous 180 days. Patients with their continuous KPCO membership <180 days prior to AF diagnosis, or aged <18 years, will be excluded. Warfarin non-users will be excluded if they die before their assigned index date (defined below), and warfarin users will be excluded if they have a purchase of warfarin during the 180 days prior to their AF diagnosis or a supply from a warfarin purchase that overlaps into the 180 days.

Patients newly diagnosed with AF from 1 January 2005 to 31 December 2008 (KPCO-I, derivation cohort) will be used to construct the prediction model cohort, while patients with a new AF diagnosis from 1 January 2009 to 31 December 2012 (KPCO-II, validation cohort) will be used as external validation of the prediction models.[31] These are both 4-year blocks of time.

Study design

We define study *start date* as the date of AF diagnosis for each patient. Study *diagnosis end date* is defined as 31 December 2008 and 31 December 2012 for the derivation and validation cohort, respectively. Study *outcome end date* is 30 June 2009 and 30 June 2013 for the derivation and validation cohort, respectively. During the time period from study start date to study outcome end date, patients who have no less than one purchase of warfarin will be categorised into *warfarin users*, while patients with no purchase of warfarin will be considered *warfarin non-users*.

Given that warfarin users may not take warfarin immediately after the diagnosis of AF (ie, study start date), there is immortal time bias in favour of warfarin.[32] That is, warfarin users who do not take warfarin initially after diagnoses as AF have to be 'immortal' before their inception of warfarin, which would provide warfarin users with an artificial survival advantage over those never on warfarin and thus overestimate the benefit of warfarin.[32 33] To control for immortal time bias, for warfarin users their study *index date* will be determined as the first date of warfarin purchase after their AF diagnosis. Subsequently, warfarin non-users will be assigned an index date that corresponds to the length of time after the study start date to the index date for their randomly matched (on year of AF diagnosis) warfarin user.[34] Warfarin non-users who die before their assigned index date will be excluded from the analysis. The length of time from study start date to the index date for warfarin users will be restricted to 180 days, in order to control the skewness of the distribution and achieve maximum matching.[34] Therefore, warfarin users whose length of time from study start day to index date exceeds 180 days will also be excluded.

All included patients will be followed up after the index date until death, termination from the KPCO system or the study outcome end date (30 June 2009 for KPCO-I, and 30 June 2013 for KPCO-II), whichever occurs first.

Outcome measures

The events of interest, including stroke, major bleeding and death, will be identified after the index date until study outcome end date. For the primary objective, all the patients will be categorised as one of the three combined outcome groups: stroke with no major bleeding, major bleeding with no stroke or neither event. For the secondary objective, patients will be grouped into either survival group or non-survival group.

Some patients may have a diagnosis of stroke and/or major bleeding predating their index date. Stroke and/or bleeding before the index date will be considered as a risk factor to reflect comorbidity (ie, prior stroke, bleeding) rather than as an end point event of interest in this study. Meanwhile, there may be some patients experiencing both a stroke and major bleeding chronologically during follow-up. We choose the event that happens first as our outcome, and categorise the patient into the corresponding group, in order to include as many outcome events as possible.

All stroke and major bleeding events will be administratively identified from the VDW Diagnosis Database recorded in an ambulatory KPCO medical office visit, ED visit or inpatient stay after the index date until outcome end date using ICD-9-CM codes. Table 1 shows the ICD-9-CM codes for stroke and major bleeding used in this study. Stroke events will be identified based on ICD-9-CM codes 433.xx, 434.xx, 436.xx in this study, in which the codes have been validated in a KPCO study with high positive predictive values.[35] Bleeding which results in an ED visit requiring a transfusion, or an admission to hospital will be considered as major bleeding.[36] We will not identify the major bleeding that causes a fall in haemoglobin level of at least 20 g/L but does not require a transfusion,[37] because no data on haemoglobin for the patients during follow-up are available in

Table 1 ICD-9-CM codes for stroke and major bleeding outcomes

Outcome measure	ICD-9-CM codes
Stroke	433.xx 434.xx 436.xx
Major bleeding	
Haemorrhagic stroke	430.xx, 431.xx, 432.xx
GI bleeding (including upper and lower GI bleeding)	
Upper GI	456.0, 530.7, 530.82; 531.0, 531.2, 531.4, 531.6; 532.0, 532.2, 532.4, 532.6; 533.0, 533.2, 533.4, 533.6; 534.0, 534.2, 534.4, 534.6; 535.01, 535.21, 535.31, 535.41, 535.51, 535.61; 537.83
Lower GI	562.02, 562.03, 562.12, 562.13; 568.81, 569.3, 569.85, 578
Other major bleeding	287.8, 287.9; 459.0, 596.7, 599.7, 627.1, 719.1, 784.7, 784.8, 786.3

GI bleeding, gastrointestinal bleeding; ICD-9-CM, International Classification of Disease, Ninth Revision, Clinical Modification diagnosis.

this study. All ICH will be counted as major bleeding rather than strokes. Information on deaths for included patients will be obtained from the VDW Death Database.

Independent variables

All patients' demographic factors, laboratory data, comorbidities, warfarin medication data and concurrent use of medication will be retrieved from the administrative KPCO databases and patients' electronic medical record. The KPCO databases link patients' pharmacy profiles, such that patients' medications information including warfarin-related data can be accessed. Approximately 94% of KPCO prescriptions are purchased at in-house pharmacies.

Specifically, in this study, demographic data include patients' gender and baseline age. Laboratory data include INR, haemoglobin, serum creatinine and albumin, where all the measures are most proximal to but before the index date.

For comorbidities, data recorded in an ambulatory KPCO medical office visit in the 180 days prior to the index date will be obtained. We will retrieve the comorbidity data including the components of the CHA_2DS_2-VASc and HAS-BLED schemes, as well as the comorbidities included in the Charlson Comorbidity Index,[38] in order to obtain a comprehensive list of comorbidities as potential independent variables. Concretely, data on comorbidities include the presence of congestive heart failure, hypertension, diabetes, prior stroke/transient ischemic attack, myocardial infarction, peripheral vascular disease, renal disease, liver disease, prior major bleeding (including GI bleeding and ICH), concomitant use of antiplatelets or non-steroidal anti-inflammatory drugs (NSAIDs), alcohol abuse, other cerebrovascular disease, dementia, peptic ulcer disease, chronic pulmonary disease, rheumatic disease, AIDS, hemiplegia or paraplegia, and any malignancy (including lymphoma and leukaemia and metastatic solid tumour, except malignant neoplasm of skin). Table 2 shows the ICD-9-CM codes for the comorbidities analysed in this study.

Warfarin medication data include the presence of sold prescriptions for warfarin from a KPCO pharmacy between the index date and outcome end date. Length of time in days from study start date to the purchase date will be recorded. In addition, data on each sold warfarin prescription's length of time from index date and days of medication supplied will be obtained.

For concurrent medication data, the presence of a sold prescription for the purchases during the 90 days after the index date will be recorded. Our analysis will be restricted to those concomitant medications for which there is evidence of an interaction that potentiates or inhibits the effect of warfarin, based on the systematic review of interactions of warfarin with other drugs.[39 40] Table 3 displays the complete medication list used in this study, which includes other anticoagulants, antiplatelets, NSAIDs and selected representatives of other families including anti-infective agents, cardiac drugs, central nervous system drugs, GI drugs, etc.

Statistical analyses

All data will be examined on a descriptive basis and presented as the mean±SD for continuous variables, and frequency and percentages for categorical variables. Student t test will be used to compare continuous variables, and χ^2 test will be applied for categorical variables between warfarin users and non-users. Unless otherwise specified, all statistical tests will be two-sided using an α level of 0.05.

Since there are three multinomial levels for the primary outcomes (stroke with no major bleeding, major bleeding with no stroke or neither event), we will use polytomous logistic regression (PLR) to develop a prediction model for patients' individual combined benefit and harm outcomes. Using neither event group as the reference group, two models will be constructed to predict stroke with no major bleeding, and major bleeding with no stroke, respectively. As regards death, we will use Cox proportional hazards regression analysis to build a prediction model for all-cause mortality.

Sample size

For logistic regression, a fitted model is likely to be reliable and stable when the number of participants with the outcome (ie, either stroke with no bleeding, or bleeding with no stroke, or bleeding with stroke) is 10–15 times the number of predictor variables.[41 42] We anticipate that about 10 predictors will be included into the PLR model maximally; therefore 100 stroke with no major bleeding, and 100 major bleeding with no stroke will be required to devise the PLR models in the derivation cohort.

Face validity of the KPCO data

We will show the trend of the incidence rates of stroke in KPCO-I and KPCO-II stratified by CHA_2DS_2-VASc score, and the incidence rates of major bleeding stratified by HAS-BLED score, which can judge the validity of the KPCO cohort for patients with AF by ensuring that it conforms to previously validated clinical prediction rules.

Model construction

For primary outcomes in patients with AF in the KPCO-I cohort, first, because predictors that are highly correlated with others contribute little independent information, pruning candidate predictors is required.[43] The effect of multicollinearity between predictors would inflate the values of the SEs of the coefficients in the model, which may drive some predictors away from statistical significance. To avoid this, the variance inflation factor with a threshold of 4 will be chosen to determine whether predictors are redundant and highly correlated.[44] Subsequently, for each pair of patients (ie, stroke with no major bleeding vs neither event, major bleeding

Table 2 ICD-9-CM codes for comorbidities

Comorbidity	ICD-9-CM codes
Components of CHA_2DS_2-VASc	
Congestive heart failure	398.91, 402.01, 402.11, 402.91, 428.x, 518.4; 404.01, 404.03, 404.11, 404.13, 404.91, 404.93
Hypertension	401.x, 402, 405
Diabetes	250, 253.5, 271.4, 275.0, 357.2, 362.0, 588.1, 648.0, 790.2, 790.6
Prior stroke/TIA	V12.54; 433.xx, 434.xx, 436.xx; 435.xx
Vascular diseases	
Myocardial infarction	410.x, 412.x
Peripheral vascular disease	093.0, 437.3, 440.x, 441.x, 443.1–443.9, 47.1, 557.1, 557.9, V43.4
Components of HAS-BLED	
Hypertension	Same as above
Abnormal renal/liver function	
Renal disease	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 585.3–585.9, 586, V42.0, V45.1, V56.x
Moderate/severe liver disease	456.0–456.2, 570, 572.2–572.8, 573.4–573.5, V42.7
Prior stroke/TIA	Same codes as stroke outcome
Prior bleeding	
Major bleeding history	Same codes as major bleeding outcome
Bleeding predisposition (anaemia)	280.8–280.9, 281.0–281.9, 282.2, 282.3, 282.8, 282.9, 283.0, 283.10, 283.19, 283.9, 284.x, 285.x, 648.2, V18.2, V78.0, V78.1
Drugs/alcohol concomitantly	
Use of antiplatelets and NSIADs	V58.63, V58.64
Alcohol abuse	291 303.00, 303.01, 303.02, 303.90, 303.91, 303.92, 305.00, 305.01, 305.02, 357.5, 425.5, 535.3
Other comorbidities	
Other cerebrovascular disease	362.34, 437.xx, 438.xx
Dementia	290.x, 294.1, 331.2
Chronic pulmonary disease	416.8, 416.9, 490.x–505.x, 506.4, 508.1, 508.8
Rheumatic disease	446.5, 710.0–710.4, 714.0– 714.2, 714.8, 725.x
Peptic ulcer disease	531.x–534.x
Hemiplegia or paraplegia	334.1, 342.x, 343.x, 344.0–344.6, 344.9
Any malignancy, including lymphoma and leukaemia and	140.x–172.x, 174.x–195.8, 200.x–208.x, 238.6
metastatic solid tumour, except malignant neoplasm of skin	196.x–199.x
AIDS/HIV	042.x–044.x

CHA_2DS_2-VASc, *C*ongestive heart failure; *H*ypertension; *A*ge ≥75 years (2 points); *D*iabetes mellitus; *S*troke (2 points); *V*ascular disease, *A*ge 65–74 years and *S*ex (female); HAS-BLED, *H*ypertension; *A*bnormal renal/liver function; *S*troke; *B*leeding history or predisposition; *L*abile international normalised ratio; *E*lderly (>65 years); *D*rugs/alcohol concomitantly; ICD-9-CM, International Classification of Disease, Ninth Revision, Clinical Modification diagnosis; NSIADs, non-steroidal anti-inflammatory drugs; TIA, transient ischaemic attack.

with no stroke vs neither event), univariate logistic regression will be performed first before selecting significant variables for multivariable regression, where the α level of 0.20 will be chosen to ensure χ^2 measures include all possible predictors. After taking multicollinearity into account and selecting significant variables based on univariate logistic regression, PLR will be used to build the prediction models.

To investigate whether warfarin can modify the effect of other predictors in the PLR models on stroke and major bleeding, all the two-way interactions between warfarin status (ie, with or without warfarin) and other predictors will be tested. Significant interactions with a priori α value of ≤0.05 between warfarin and other predictors will be retained and added into the prediction models. Moreover, given the importance and potential interactions of the predictors composing the CHA_2DS_2-VASc and HAS-BLED schemes (eg, sex, age, hypertension), we will also evaluate the two-way interactions along with their main effect terms if they are kept in our PLR models.[45] Then the significant interactions with an α value of ≤0.05 will be included to update and finalise the PLR models. For instance, if hypertension and age are included in our PLR models, we will also test the significance of their two-way interaction (hypertension×age) before choosing this interaction into the finalised PLR models.

For secondary outcome, Cox regression model will be applied to building the prediction model of death. Similar procedures to those for primary outcomes will

Table 3 Concurrent medication list* included for analysis

Concurrent medication group	Drug list
Other anticoagulants	Dalteparin, fondaparinux, heparin, tinzaparin, apixaban, dabigatran, rivaroxaban
Antiplatelets	Abciximab, Aggrenox (dipyridamole+ASA), aspirin, clopidogrel, dipyridamole, eptifibatide, prasugrel, ticagrelor, ticlopidine, tirofiban
NSAIDs	Celecoxib, diclofenac, etodolac, fenoprofen, flurbiprofen, ibuprofen, indomethacin, ketoprofen, ketorolac, mefenamic acid, meloxicam, nabumetone, naproxen, oxaprozin, piroxicam, rofecoxib, sulindac, tolmetin, valdecoxib
Antibiotics	Amoxicillin, amoxycillin/clavulanic acid, nafcillin, cefaclor, cefadroxil, cefazolin, cefixime, cefoxitin, cefprozil, cefradine, ceftriaxone, cefuroxime, cephalexin, ciprofloxacin, levofloxacin, moxifloxacin, doxycycline, erythromycin, cotrimoxazole, rifampin
Antifungals	Fluconazole, metronidazole, miconazole, voriconazole, griseofulvin
Antitubercular agents	Isoniazid
Cardiac drugs	Amiodarone, propafenone, diltiazem, propranolol
Antilipemic drugs	Clofibrate, fenofibrate, cholestyramine
Antidepressants	Citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline
Other CNS drugs	Entacapone, carbamazepine, Butalbital, pentobarbital, Phenobarbital, thiopental
GI drugs	Cimetidine, omeprazole, lansoprazole, esomeprazole, pantoprazole, rabeprazole
Other drugs	Tramadol

*Only the medications with evidence of an interaction that potentiates or inhibits the effect of warfarin are included.
ASA, acetylsalicylic acid; CNS drugs, central nervous system drugs; GI drugs, gastrointestinal drugs; NSAIDs, non-steroidal anti-inflammatory drugs.

be followed, that is, choose the variables without multicollinearity and those significant predictors in univariate analysis to make up the model, then include the significant interactions (warfarin×other predictors, two-way interactions of the predictors composing the CHA_2DS_2-VASc and HAS-BLED schemes) to finalise the model to predict death. A statistical test of proportional hazards assumption and a graphical examination using Schoenfeld residuals will be carried out to assess the proportional hazards assumption.[46]

Sensitivity analyses

For missing data, if <10% of observations on a variable are missing, the mean or median of the variable in its group will be used for imputation. If no less than 10% of data are missing, assuming they are missing at random, multiple imputations will be performed using clinical judgement to identify factors to be included in the imputation model.[47 48] If multiple imputations are used, as a sensitivity analysis, the obtained PLR results will be compared with the original PLR models with missing data.

Since there may be gaps in the consumption of warfarin for the patients during follow-up, another sensitivity analysis will be conducted using warfarin as a time-dependent covariate, to investigate whether the effect of warfarin is robust on stroke and major bleeding in the PLR model, and death in the Cox model, respectively.[49] We will use the gap of >30 days between the last day when a previous purchase is expected to run out and the first day of the next purchase, to indicate warfarin discontinuation for warfarin users.

Moreover, to investigate whether the predictors are sensitive after taking all-cause death (as a competing risk of stroke and major bleeding) into account, we will perform a competing risk analysis to obtain the hazard functions for stroke and major bleeding separately. Two proportional subdistribution hazards models of the Fine and Gray method[50] will be constructed for stroke and major bleeding, respectively. In the competing risk analysis, patients who die ahead of an event of stroke or major bleeding will be left in the risk set with decreasing weight to account for declining observability, rather than being treated as simple censoring.[50] Predictors and their coefficients in the proportional subdistribution hazards models will be used to compare with those in the PLR models.

Model performance

To assess calibration of the PLR models for primary outcomes, we will compare the predicted risks of stroke with no major bleeding, and major bleeding with no stroke to the observed event rates in different deciles of predicted risks.[51 52] Differences between predicted and observed event rates will be used to calculate a Hosmer-Lemeshow statistic, where a non-significant result indicates no evidence of lack of fit to the data. To assess discriminability, we will calculate the area under the two receiver operating characteristic curves (AUC) for each pair of comparison: stroke with no major bleeding versus neither event, and major bleeding with no stroke versus neither event.

Regarding the Cox model for all-cause death, we will evaluate the model calibration by comparing the predicted risk of death and observed rates across each 10th of the observed risk,[51 52] where the observed risk will be calculated using the Kaplan-Meier product-limit estimate. Goodness-of-fit of the model will be investigated

using a Gronnesby and Borgan test with 10 groups according to the predicted risk score, in which a non-significant result implies the model is a good fit.[53] Harrell's C index will be calculated to assess the discriminability of the model.[54]

For a typical newly diagnosed patient with AF, we will input his/her individual information into the PLR models and calculate the probability of stroke with no major bleeding, major bleeding with no stroke and neither event, respectively.[55] We can also calculate his/her probability of death using the Cox regression model.

Model validation

As internal validation of the PLR models, a 10-fold cross-validation and a bootstrap analysis resampling 1000 times with replacement will be conducted to assess the models' validation. The AUCs of the original PLR models will be compared with those of cross-validation, while coefficients from the original PLR models will be contrasted with those from bootstrap models.

KPCO-II cohort will be used for external validation of the PLR models. Because the incidences of stroke and major bleeding in KPCO-I and KPCO-II cohort may be different, we will update the original models for the validation cohort.[55–57] Then the assessment of calibration, goodness-of-fit and discriminability will be again performed in KPCO-II cohort. We will also use KPCO-II cohort to externally validate the prediction model for death.

Ethics and dissemination

Results from this study will be published in a peer-reviewed journal electronically and in print. The prediction models can provide comprehensive information on the individual combined benefit and harm with and without warfarin for patients with AF, which may aid in patient-physician shared decision-making when they are considering warfarin therapy. For the warfarin users, the models will also help enhance patients' medication adherence once the patients are clear about their individual predicted risk of outcomes, after they initialise warfarin therapy in the real world.

Author affiliations
[1]Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Ontario, Canada
[2]St. Joseph's Hospital, McMaster University, Hamilton, Ontario, Canada
[3]Department of Medicine, McMaster University, Hamilton, Ontario, Canada
[4]Kaiser Permanente Colorado Clinical Pharmacy Research Team, Aurora, Colorado, USA
[5]University of Colorado Skaggs School of Pharmacy and Pharmaceutical Sciences, Denver, Colorado, USA
[6]Department of Pharmacotherapy, University of Utah, Salt Lake City, Utah, USA

Acknowledgements We thank Dr Jennifer A Pereira for her substantial preparatory contribution in her unpublished doctoral thesis (JA Pereira. Unpublished doctoral dissertation, 2008) to this study.

Contributors GL, AH, TD, DMW and LT were responsible for the study conception and design. GL, AH and LT were responsible for the drafting of the manuscript. AH, TD, DMW, MAHL and LT made critical revisions and provided professional and statistical support. All authors approved the final version of the manuscript.

Funding GL received a Father Sean O'Sullivan Research Award, the Research Institute of St. Joe's Hamilton, and a doctoral award from the CSC.

Competing interests None declared.

Ethics approval the KPCO Institutional Review Board (No.: CO-14-2093) and the Hamilton Integrated Research Ethics Board (No.: 14–394-C).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Hart RG, Halperin JL, Pearce LA, *et al*. Lessons from the Stroke Prevention in Atrial Fibrillation trials. *Ann Intern Med* 2003;138:831–8.
2. Rockson SG, Albers GW. Comparing the guidelines: anticoagulation therapy to optimize stroke prevention in patients with atrial fibrillation. *J Am Coll Cardiol* 2004;43:929–35.
3. Camm AJ, Kirchhof P, Lip GY, *et al*. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur Heart J* 2010;31:2369–429.
4. Singer DE, Albers GW, Dalen JE, *et al*. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 2008;133:546S–92S.
5. Cairns JA, Connolly S, McMurtry S, *et al*. Canadian Cardiovascular Society atrial fibrillation guidelines 2010: prevention of stroke and systemic thromboembolism in atrial fibrillation and flutter. *Can J Cardiol* 2011;27:74–90.
6. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med* 2007;146:857–67.
7. Ghanny S, Crowther M. Treatment with novel oral anticoagulants: indications, efficacy and risks. *Curr Opin Hematol* 2013;20:430–6.
8. Wells G, Coyle D, Cameron C, *et al*. Safety, effectiveness, and cost-effectiveness of new oral anticoagulants compared with warfarin in preventing stroke and other cardiovascular events in patients with atrial fibrillation [Internet]. Ottawa: Canadian Agency for Drugs and Technologies in Health, 2012:173. (CADTH Therapeutic Review). http://www.cadth.ca/media/pdf/NOAC_Therapeutic_Review_final_report.pdf
9. Go AS, Hylek EM, Chang Y, *et al*. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? *JAMA* 2003;290:2685–92.
10. Flaherty ML, Kissela B, Woo D, *et al*. The increasing incidence of anticoagulant-associated intracerebral hemorrhage. *Neurology* 2007;68:116–21.
11. Holbrook A, Labiris R, Goldsmith CH, *et al*. Influence of decision aids on patient preferences for anticoagulant therapy: a randomized trial. *CMAJ* 2007;176:1583–7.
12. Ansell J. Warfarin versus new agents: interpreting the data. *Hematology Am Soc Hematol Educ Program* 2010;2010:221–8.
13. Arepally G, Bauer KA, Bhatt DL, *et al*. The use of antithrombotic therapies in the prevention and treatment of arterial and venous thrombosis: a survey of current knowledge and practice supporting the need for clinical education. *Crit Pathw Cardiol* 2010;9:41–8.
14. McCrory DC, Matchar DB, Samsa G, *et al*. Physician attitudes about anticoagulation for nonvalvular atrial fibrillation in the elderly. *Arch Intern Med* 1995;155:277–81.
15. Holbrook A, Pullenayegum E, Troyan S, *et al*. Personalized benefit-harm information influences patient decisions regarding warfarin. *J Popul Ther Clin Pharmacol* 2013;20:e406–15.
16. Lip GY, Nieuwlaat R, Pisters R, *et al*. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* 2010;137:263–72.

17. Gage BF, Waterman AD, Shannon W, *et al*. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA* 2001;285:2864–70.
18. Pisters R, Lane DA, Nieuwlaat R, *et al*. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest* 2010;138:1093–100.
19. Lip GY. Implications of the CHA(2)DS(2)-VASc and HAS-BLED Scores for thromboprophylaxis in atrial fibrillation. *Am J Med* 2011;124:111–14.
20. Olesen JB, Lip GY, Hansen PR, *et al*. Bleeding risk in 'real world' patients with atrial fibrillation: comparison of two established bleeding prediction schemes in a nationwide cohort. *J Thromb Haemost* 2011;9:1460–7.
21. Lip GY, Frison L, Halperin JL, *et al*. 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, Drugs/Alcohol Concomitantly) score. *J Am Coll Cardiol* 2011;57:173–80.
22. Singer DE, Chang Y, Fang MC, *et al*. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. *Ann Intern Med* 2009;151:297–305.
23. Friberg L, Rosenqvist M, Lip GY. Net clinical benefit of warfarin in patients with atrial fibrillation: a report from the Swedish atrial fibrillation cohort study. *Circulation* 2012;125:2298–307.
24. Olesen JB, Lip GY, Lindhardsen J, *et al*. Risks of thromboembolism and bleeding with thromboprophylaxis in patients with atrial fibrillation: a net clinical benefit analysis using a 'real world' nationwide cohort study. *Thromb Haemost* 2011;106:739–49.
25. Lip GY, Tse HF, Lane DA. Atrial fibrillation. *Lancet* 2012;379:648–61.
26. Devereaux PJ, Anderson DR, Gardner MJ, *et al*. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ* 2001;323:1218.
27. MacLean S, Mulla S, Akl EA, *et al*. Patient values and preferences in decision making for antithrombotic therapy: a systematic review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2012;141(2 Suppl):e1S–23S.
28. Miyasaka Y, Barnes ME, Bailey KR, *et al*. Mortality trends in patients diagnosed with first atrial fibrillation: a 21-year community-based study. *J Am Coll Cardiol* 2007;49:986–92.
29. Gage BF, Yan Y, Milligan PE, *et al*. Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF). *Am Heart J* 2006;151:713–19.
30. Witt DM, Sadler MA, Shanahan RL, *et al*. Effect of a centralized clinical pharmacy anticoagulation service on the outcomes of anticoagulation therapy. *Chest* 2005;127:1515–22.
31. Altman DG, Vergouwe Y, Royston P, *et al*. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
32. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol* 2008;167:492–9.
33. Lévesque LE, Hanley JA, Kezouh A, *et al*. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ* 2010;340:b5087.
34. Zhou Z, Rahme E, Abrahamowicz M, *et al*. Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods. *Am J Epidemiol* 2005;162:1016–23.
35. Olson KL, Wood MD, Delate T, *et al*. Positive predictive values of ICD-9 codes to identify patients with stroke or TIA. *Am J Manag Care* 2014;20:e27–34.
36. Witt DM, Delate T, Clark NP, *et al*. Outcomes and predictors of very stable INR control during chronic anticoagulation therapy. *Blood* 2009;114:952–6.
37. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost* 2005;3:692–4.
38. Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
39. Holbrook AM, Pereira JA, Labiris R, *et al*. Systematic overview of warfarin and its drug and food interactions. *Arch Intern Med* 2005;165:1095–106.
40. Holbrook A, Schulman S, Witt DM, *et al*. Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2012;141: e152S.
41. Peduzzi P, Concato J, Kemper E, *et al*. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.
42. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis (Springer series in statistics)*. Springer, 2001.
43. Royston P, Moons K, Altman DG, *et al*. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009;338:b604.
44. Miles J, Shevlin M. *Appling regression and correlation: a guide for students*. London: Sage Publications Limited, 2001.
45. McWilliams A, Tammemagi MC, Mayo JR, *et al*. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;369:910–19.
46. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994;81:515–26.
47. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat* 2007;61:79–90.
48. Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol* 2009;60:549–76.
49. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health* 1999;20:145–57.
50. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999;94:496–509.
51. Wang TJ, Massaro JM, Levy D, *et al*. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. *JAMA* 2003;290: 1049–56.
52. Nijman RG, Vergouwe Y, Thompson M, *et al*. Clinical prediction model to aid emergency doctors managing febrile children at risk of serious bacterial infections: diagnostic study. *BMJ* 2013;346:f1706.
53. Gronnesby JK, Borgan O. A method for checking regression models in survival analysis based on the risk score. *Lifetime Data Anal* 1996;2:315–28.
54. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
55. Wijesinha A, Begg CB, Funkenstein HH, *et al*. Methodology for the differential diagnosis of a complex data set. A case study using data from routine CT scan examinations. *Med Decis Making* 1983;3:133–54.
56. Moons KG, Kengne AP, Grobbee DE, *et al*. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98:691–8.
57. Janssen KJ, Moons KG, Kalkman CJ, *et al*. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008;61:76–86.

Chapter 3

Can we predict individual combined benefit and harm of therapy? Warfarin therapy for atrial fibrillation as a test case

Running title: prediction of individual combined benefit and harm of therapy

Guowei Li[1,2], Lehana Thabane[1,2,*], Thomas Delate[3,4], Daniel M. Witt[5], Mitchell A.H. Levine[1,2,6], Ji Cheng[1,2], Anne Holbrook[1,2,6,*]

[1] Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, ON, Canada

[2] St. Joseph's Hospital, McMaster University, Hamilton, ON, Canada

[3] Kaiser Permanente Colorado Clinical Pharmacy Research Team, Aurora, CO, USA

[4] University of Colorado Skaggs School of Pharmacy and Pharmaceutical Sciences, Denver, CO, USA

[5] Department of Pharmacotherapy, University of Utah, Salt Lake City, Utah, USA

[6] Division of Clinical Pharmacology & Toxicology, Department of Medicine, McMaster University, Hamilton, ON, Canada

*Corresponding authors

Anne Holbrook, MD, PharmD, MSc

Director/Professor, Division of Clinical Pharmacology and Toxicology, Department of Medicine, McMaster University

SJHH G624, 50 Charlton Avenue E, Hamilton ON, L8N 4A6

E-mail: holbrook@mcmaster.ca

Telephone: 905-522-1155 x 35269; Fax: 905-540-6520

and

Lehana Thabane, PhD, MSc

Professor/Associate Chair, Clinical Epidemiology and Biostatistics, McMaster University

Father Sean O'Sullivan Research Centre, St. Joseph's Healthcare Hamilton

3rd Floor Martha, Room H325, 50 Charlton Avenue E, Hamilton ON, L8N 4A6

E-mail: thabanl@mcmaster.ca

Telephone: 905-522-1155 x 33720; Fax: 905-308-7386

Abstract

Objectives: To construct and validate a prediction model for individual combined benefit and harm outcomes (stroke with no major bleeding, major bleeding with no stroke, neither event, or both) in patients with atrial fibrillation (AF) with and without warfarin therapy.

Methods: Using the Kaiser Permanente Colorado databases, we included patients newly diagnosed with AF between January 1, 2005 and December 31, 2012 for model construction and validation. The primary outcome was a prediction model of composite of stroke or major bleeding using polytomous logistic regression (PLR) modelling. The secondary outcome was a prediction model of all-cause mortality using the Cox regression modelling.

Results: We included 9074 patients with 4537 and 4537 warfarin users and non-users, respectively. In the derivation cohort (n = 4632), there were 136 strokes (2.94%), 280 major bleedings (6.04%) and 1194 deaths (25.78%) occurred. In the prediction models, warfarin use was not significantly associated with risk of stroke, but increased the risk of major bleeding and decreased the risk of death. Both the PLR and Cox models were robust, internally and externally validated, and with acceptable model performances.

Conclusions: In this study, we introduce a new methodology for predicting individual combined benefit and harm outcomes associated with warfarin therapy for patients with AF. Should this approach be validated in other patient populations, it has potential advantages over existing risk stratification approaches as a patient-physician aid for shared decision-making

Keywords: atrial fibrillation; warfarin; prediction model; polytomous logistic regression; stroke; major bleeding

Introduction

Atrial fibrillation (AF) is a common, age-related, chronic arrhythmia that is a major risk factor for stroke and mortality [1,2]. At present, oral anticoagulants are the mainstay for stroke prophylaxis in patients with AF [3]. Despite the growth in use of newer oral anticoagulants, warfarin remains a dominant antithrombotic therapy for AF, where it lowers rates of stroke as well as mortality [2-5]. However, the use of anticoagulants also is associated with an increased risk of major bleeding including intracranial hemorrhage (ICH). Thus, this combination of potential life-saving benefit and life-threatening harm may dissuade clinicians from prescribing warfarin for eligible patients [6-10].

Clinical prediction rules such as the $CHADS_2$ (**C**ongestive heart failure, **H**ypertension, **A**ge > 75 years, **D**iabetes, Previous **s**troke [2 points]) and the CHA_2DS_2-VASc (**C**ongestive heart failure; **H**ypertension; **A**ge ≥ 75 years [2 points]; **D**iabetes mellitus; **S**troke [2 points], **V**ascular disease, **A**ge 65-74 years, and **S**ex category [female]) scores have been developed and widely used to predict stroke risk in AF patients [2,4,11,12]. Likewise, the HAS-BLED score (**H**ypertension; **A**bnormal renal/liver function; **S**troke history; **B**leeding history or predisposition; **L**abile international normalized ratio [INR], **E**lderly [>65 years]; **D**rugs/alcohol concomitantly) has been validated to predict risk of major bleeding with warfarin therapy [2,4,13-16]. Unfortunately, the $CHADS_2$, CHA_2DS_2-VASc and HAS-BLED scores were not devised in the same population and thus are unable to assess simultaneously a patient's potential for benefit and/or harm with warfarin therapy [17].

While the $CHADS_2$, CHA_2DS_2-VASc and HAS-BLED scores help estimate an individual's chance of benefit and harm separately, a more sophisticated methodology is needed. The 'net benefit' approach involves calculating the main benefit of warfarin therapy (reduced risk of stroke or systemic embolism) then deducting the main harm (weight*increased risk of ICH, weight = 1.5) in the same population [18-20]. However, this approach does not take into account gastrointestinal (GI) bleeding risk, and the weighting for ICH is chosen arbitrarily.

In general, treatment effects of warfarin therapy for individual patients can be divided into four quadrants: 1) benefit without harm; 2) harm without benefit; 3) neither benefit or harm; and 4) both benefit and harm simultaneously (**Table 1**). A method for predicting the probabilities of the four outcome quadrants (i.e., individualized combined benefit and harm outcomes) for each patient is needed. The polytomous logistic regression (PLR) modelling can be used for predictions due to the four multinomial levels of outcomes [21,22]. Therefore, the objective of this study was to use the PLR modelling to construct and externally validate a prediction model for patients' individual combined benefit and harm outcomes (stroke with no major bleeding, major bleeding with no stroke, neither event, or both stroke and major bleeding) with and without warfarin therapy for AF. In real-world clinical settings, the prediction of individualized combined benefit and harm outcomes related to warfarin therapy could assist with the patient-physician shared decision-making process.

Methods

Study design and setting

The methods have been described in detail previously [17]. Briefly, Kaiser Permanente Colorado (KPCO), a non-profit, integrated health care delivery system in the U.S. Denver-Boulder metropolitan area, utilizes a centralized anticoagulation service that provides anticoagulation services for KPCO patients with AF [23,24]. KPCO maintains extensive medical, pharmacy, laboratory, utilization, mortality, and membership electronic, integrated administrative datasets. Data were extracted for KPCO patients diagnosed with AF who were and were not prescribed warfarin therapy and analyzed at St. Joseph's Healthcare Hamilton in Hamilton, ON. The KPCO Institutional Review Board and the Hamilton Integrated Research Ethics Board approved this study with a waiver for informed consent.

Patients newly diagnosed with AF between January 1, 2005 and December 31, 2012 were included. Newly diagnosed status was defined by absence of AF diagnosis in the previous 180 days. Patients were followed for up to 180 days after AF diagnosis to assess if warfarin therapy was initiated. Patients who had at least one warfarin purchase or no warfarin purchases were grouped as warfarin *users* and *non-users*, respectively. Warfarin non-users were randomly

matched 1:1 to warfarin users on year of AF diagnosis [25]. Patients with AF diagnosed between January 1, 2005 and December 31, 2008 comprised the derivation cohort (KPCO-I), while patients with AF diagnosed between January 1, 2009 and December 31, 2012 comprised the validation cohort (KPCO-II) [26].

Study patients

The date of AF diagnosis for each patient was defined as study ***start date***. To include as many outcomes as possible, study ***outcome end date*** was defined as June 30, 2009 and June 30, 2013 for the derivation and validation cohorts, respectively. To control the potential of immortal time bias, the study ***index date*** for warfarin users was defined as the first warfarin purchase date after start date [27,28]. Warfarin non-users were assigned an index date corresponding to the length of time from study start date to the index date of their randomly-matched warfarin user [25]. A warfarin non-user who died before her/his assigned index date and her/his matched user were excluded. Patients were followed from index date until KPCO plan disenrollment, death, or study outcome end date, whichever came first [17].

Outcomes

The primary outcome was a prediction model of composite of stroke or major bleeding. The secondary outcome was a prediction model of all-cause death. All of the outcomes were assessed from the index date to outcome end date. For the prediction model of stroke or major bleeding, we categorized patients into one of the four outcome groups based on their survival time to first event: stroke with no major bleeding, major bleeding with no stroke, neither event, or both stroke and major bleeding. For the prediction model of all-cause mortality, patients were categorized into survival or non-survival groups.

Stroke and major bleeding events were identified during an ambulatory KPCO medical office visit, emergency department (ED) visit, or inpatient stay using International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) codes in the primary position. Major bleeding was defined as bleeding that led to a hospital admission or an ED visit requiring a transfusion [29]. However, bleeding that caused a drop in hemoglobin

of ≥ 20g/L but did not necessitate a transfusion [30] was not included as major bleeding since no inpatient or ED hemoglobin laboratory values were available. ICH was categorized as major bleeding, rather than stroke. Stroke or major bleeding occurring before the index date was categorized as a risk factor (i.e., prior stroke, prior major bleeding) rather than a study outcome [17].

Potential predictors of benefit and/or harm

The potential predictors used in this study included patients' demographic characteristics (e.g., sex, age), laboratory measures, baseline comorbidities, warfarin use, and concurrent use of medications that interact with warfarin. Laboratory measurements included INR, hemoglobin, serum creatinine and albumin recorded most proximal but prior to the index date. Comorbidities were from ambulatory KPCO medical office visits in the 180 days prior to the index date. Comorbidities were components of the CHA_2DS_2-VASc and HAS-BLED schemes, as well as components included in the Charlson Comorbidity Index [31]. Data on warfarin use included the length of time from study start date to the first purchase date, the length of time for each dispensed warfarin prescription from index date, and days of warfarin supplied. Concurrent use of other medications included purchases for medications made during the 90 days after index date. Medications for which there was evidence of an interaction that inhibited or potentiated the effect of warfarin, according to the findings from systematic reviews [32,33], were included.

Statistical analyses

All tests were two-sided with a significance level of 0.05, unless otherwise specified. We described continuous variables as means (+/- standard deviations [SDs]), and frequencies and percentages for categorical variables. Student's t-tests were used to compare continuous variables and chi-square tests of associations were applied for categorical variables. In the derivation and validation cohort, we assessed the stroke and major bleeding incidence rate trends stratified by the CHA_2DS_2-VASc score and HAS-BLED score, respectively.

Model building

PLR modeling was used to develop a prediction model for the four individual benefit and harm outcomes using the neither event group as the referent category. Odds ratios (ORs) with 95% confidence intervals (CIs) were used to quantify the relationship between outcomes and predictors. We employed Cox proportional hazards regression analysis to build a prediction model for all-cause mortality, using hazard ratios (HRs) to quantify the associations between predictors and mortality.

Both the PLR and Cox regression models followed the same procedures for model construction. First, the effect of multicollinearity was evaluated using the criterion of a variance inflation factor ≥ 4 to prune candidate predictors. Subsequently, we performed univariate analyses to select significant variables for multivariable analyses with an alpha level ≤ 0.20. Lastly we identified significant two-way interactions to finalize our prediction models [17].

For the primary outcome, three sensitivity analyses were performed by: 1) using multiple imputations if missing data were $\geq 10\%$; 2) treating the use of warfarin as a time-dependent covariate to evaluate the effect of warfarin on stroke and major bleeding, using a gap of > 30 days to indicate warfarin discontinuation [34]; and 3) employing a competing risk analysis using the Fine and Gray method to take into account all-cause mortality as a competing risk of stroke and major bleeding [35].

Model performance and validation

Comparison between the predicted and observed risks in deciles was used to evaluate calibration of the prediction models. Discrimination was measured by the area under the receiver operating characteristic curves (AUCs) for the PLR model and Harrell's C index for the Cox model. Goodness-of-fit was assessed by a Hosmer-Lemeshow statistic and Gronnesby and Borgan test with ten groups based on the predicted risk scores for the PLR and Cox models, respectively [17].

Two internal validations were performed for the PLR model by using 10-fold

cross-validation and bootstrap analysis. We also used bootstrap analysis to internally validate the Cox model for all-cause mortality. For the external validation, because the incidence rates of outcomes were different from the derivation and validation cohorts and there was evidence that the original models were not a good fit to the validation cohort, we updated the models' intercepts as well as the regression coefficients by using the calibration intercepts and calibration slopes [22,36,37]. The evaluation of goodness-of-fit, calibration, and discrimination was repeated in the validation cohort.

Analyses were performed with the software packages SAS Version 9.3 (SAS Institute, Inc., Cary, NC) and STATA Version 12 (Stata Corp., College Station, TX, USA). For the calibration plots of the PLR model, we used the software R version 3.2.1 (R Foundation for Statistical Computing, Vienna, Austria) with the Design library.

Results

Patient characteristics

We included 9074 patients diagnosed with AF with 4537 and 4537 warfarin users and non-users, respectively (see **Supplemental Figure 1** for patient dispositions). Overall mean age was 71.7 years (SD: 13.0) and 46% were female (**Table 2**). Overall mean CHA_2DS_2-VASc and HAS-BLED scores were 2.99 (SD: 1.56) and 1.73 (SD: 0.88), respectively.

The derivation cohort (KPCO-I) included 4632 patients with a median follow-up of 652 days, while the validation cohort (KPCO-II) included 4442 patients with a median follow-up of 628 days (**Table 2**). In the KPCO-I cohort, warfarin users were significantly older and had higher proportions of patients with congestive heart failure, hypertension, renal disease, prior major bleeding, anemia, and alcohol abuse than non-users (all $p < 0.05$). The CHA_2DS_2-VASc (mean 3.09 versus 2.73) and HAS-BLED (mean 1.80 versus 1.54) scores were higher in warfarin users. A higher proportion of warfarin users had purchased concurrently an NSAID, antibiotic, cardiac drug, GI drug, and other drug (tramadol) than non-users. However, a lower percentage of antiplatelet use was observed in warfarin users

compared with non-users (p = 0.001). Similar characteristics and comparison between warfarin users and non-users were found in the KPCO-II cohort (**Table 2**). **Supplemental Table 1** presents the comparison between warfarin users and non-users in the whole cohort (i.e., KPCO-I combined with KPCO-II), with similar results to findings as those from the KPCO-I cohort alone.

Twenty-eight patients (12 and 16 in the KPCO-I and KPCO- II cohorts, respectively) had a stroke and major bleeding outcome on the same date; thus, their time to first event could not be identified. Because of the low frequency, these patients were randomly allocated into either stroke with no major bleeding (n=14) or major bleeding with no stroke (n=14). Therefore, in the combined cohort there were 278 strokes (3.06%), 453 major bleedings (4.99%) and 2186 deaths (24.09%) occurred during follow-up. Of these, 136 strokes (2.94%), 280 major bleedings (6.04%) and 1194 deaths (25.78%) occurred in the KPCO-I cohort. In both the KPCO-I and KPCO- II cohorts, the rates of major bleeding and death, but not stroke, differed between warfarin users and non-users (**Table 3)**. Also, as shown in **Supplemental Figure 2,** there was a significant difference in all-cause mortality (log-rank p-value = 0.001) between the KPCO-I cohort and KPCO-II cohort.

Significant trends for increasing stroke and major bleeding rates with higher CHA_2DS_2-VASc and HAS-BLED scores were found (p < 0.001) for both the KPCO-I and KPCO-II cohorts **(Supplemental Table 2)**.

PLR Model

The PLR model included age, female sex, warfarin use, CHF, other cerebrovascular disease, hypertension, diabetes, prior major bleeding, prior stroke, renal disease, and concurrent use of antibiotics, antiplatelets, and GI drugs (**Table 4)**. Warfarin use was not associated with stroke (OR = 0.94, 95% CI: 0.66 - 1.34) but was associated with increased risk of major bleeding (OR = 1.71, 95% CI: 1.32 - 2.22). All other predictors in the model were associated with an increased risk of outcomes, except hypertension (OR = 0.88, 95% CI: 0.56 - 1.39) and antibiotic use (OR = 0.98, 95% CI: 0.64 - 1.51) for stroke, and female sex (OR = 0.73, 95%

CI: 0.56 - 0.94), hypertension (OR = 0.94, 95% CI: 0.66 - 1.33), and prior stroke (OR = 0.73, 95% CI: 0.40 - 1.34) for major bleeding.

Cox Model

The all-cause mortality model included age, warfarin, anemia, other cerebrovascular disease, CHF, diabetes, hypertension, prior major bleeding, malignancy, and concurrent use of antifungals and antidepressants (**Table 5**). Warfarin use was associated with a decreased risk of death (HR = 0.55, 95% CI: 0.49 - 0.62). All other predictors were associated with increased risk of death except hypertension (HR = 0.76, 95% CI: 0.66 - 0.85).

Sensitivity Analyses

When warfarin use was treated as a time-dependent covariate, similar associations between warfarin and outcomes were found as in the PLR model for stroke and major bleeding and the Cox model for all-cause mortality (**Supplemental Table 3**). Results from the competing risk sensitivity analysis for stroke and major bleeding identified similar coefficients for all the predictors included in the PLR model, indicating the robustness of the prediction model (**Table 6)**.

Model performance and validation

The prediction models had a good fit to the data in the derivation cohort ($p > 0.05$) (**Table 7)**. The discrimination of the models (AUC = 0.71 and 0.72 for stroke and major bleeding, respectively, and C index = 0.75 for all-cause mortality) were acceptable. The overall calibration of the PLR model (**Supplemental Figure 3 and 4**) and the Cox model (**Supplemental Figure 5**) was satisfactory. Bootstrap analyses for the PLR model and the Cox model yielded the same predictors and similar coefficients as the original models, indicating internal model validation (**Table 4 and 5**). Findings from 10-fold cross-validation also produced similar AUCs to the original PLR model: 0.69 for stroke and 0.71 for major bleeding (**Table 7**). For external validation in the KPCO-II cohort, the models' intercepts and the regression coefficients were updated (**Supplemental Table 4**). Results of the model goodness-of-fit test (**Table 7**), discrimination (**Table 7**) and calibration (**Supplemental**

Figure 6, 7 and 8) supported external validation for the PLR and Cox models.

Discussion

In this study of patients diagnosed with AF who were and were not initiated on warfarin therapy, we present a new methodology to predict individual combined benefit and harm outcomes of warfarin therapy. We utilized a PLR model to predict the individual benefit and harm outcomes due to its simplicity and flexibility, especially in predictor selection [21,22]. The PLR modelling can incorporate individual baseline characteristics of patients to estimate individual probabilities of the combined benefit and harm outcomes. Compared with the decision tree model which is another commonly-used method for prediction building, the PLR models have shown greater discrimination and predictive accuracy [38-43].

We found that warfarin use, age, female sex, CHF, other cerebrovascular disease, hypertension, diabetes, prior major bleeding, prior stroke, renal disease, and concurrent use of antibiotics, antiplatelets, and GI drugs were included in the PLR model for stroke and major bleeding. Our model performance was acceptable and robust. Using the predictors we identified, the estimated probabilities of the potential outcomes can be computed. For example, if an 82 year-old woman taking warfarin had CHF, diabetes, renal disease and prior major bleeding, and used GI medications concurrently with warfarin, then her log(stroke/neither event) would be -0.85, and log(major bleeding/neither event) would be -0.33, respectively. Subsequently, her estimated 3-year probability of stroke would be: $\frac{e^{-0.85}}{1+e^{-0.85}+e^{-0.33}}$ = 19.9%, her probability of major bleeding would be: $\frac{e^{-0.33}}{1+e^{-0.85}+e^{-0.33}}$ = 33.6%, and her probability of neither event would be: $\frac{1}{1+e^{-0.85}+e^{-0.33}}$ = 46.5% [22]. By contrast if she did not start warfarin therapy but all other factors were the same, her estimated probability of stroke, major bleeding and neither event would be 24.3%, 22.5% and 53.2%, respectively. Likewise, her estimated 3-year probability of all-cause mortality with and without warfarin therapy initiation would be 6.9% and 24.4% respectively, using the Cox model.

In our prediction models, warfarin was associated with an increased risk of major bleeding and decreased risk of death, which is in accordance with previous findings [44,45]. However, we did not identify an association between warfarin use and decreased risk of stroke. A possible explanation for this unexpected observation might include lack of INR control measures, such as time in therapeutic range (TTR), in our prediction models. Prior research indicates that the full benefit of stroke risk reduction may require an individual TTR of at least 70% in warfarin users [46]. However, individual TTR results for patients in our cohorts could not be included in the models since warfarin non-users were unmeasured on this factor. Another possible explanation relates to our use of ICD-9-CM codes alone to identify stroke and bleeding outcomes without confirmatory chart review. The positive predictive values of ICD-9-CM codes for bleeding have been shown to be higher than those for stroke [47,48]; thus, the use of ICD-9-CM codes alone may have provided a high rate of stroke false positives. In addition, a stroke history may have increased the likelihood that a given patient received warfarin to prevent further stroke risk and concurrently increased the likelihood that false positive stroke ICD-9-CM codes were identified during administrative data acquisition.

The $CHADS_2$/CHA_2DS_2-VASc, and HAS-BLED scores are used worldwide in patients with AF to stratify the risk of stroke and major bleeding, respectively. However, these risk-stratification tools cannot provide the individual combined benefit and harm assessments needed by patients and physicians when inception of warfarin therapy is under consideration. Moreover, concerns have been expressed about their scoring algorithms and poor discrimination [49-53]. For instance, in one study compared with their peers with a CHA_2DS_2-VASc score of 0 and 1 for men and women, respectively, the unequal risk of stroke for the additional risk factors resulted in different weighting in the scoring algorithm. This corresponded to a HR of from 1.68 with vascular disease to 3.09 with an age of 65-74 years for men and a HR of from 1.71 with hypertension to 3.03 with an age of 65-74 years for women [51]. Therefore given the potential different weighting for individual components of the scores as well as more detailed information provided by the individual components, we

used individual risk factors, rather than gross risk scores, in our model construction.

Other studies have used the 'net benefit' approach of considering stroke and major bleeding outcomes simultaneously [18-20]. Unfortunately, GI bleeding risk was not considered, and the weighting factor reflecting the importance of ICH was chosen subjectively and arbitrarily in these studies. Additionally, while some studies have combined stroke and bleeding risk-stratification scores to calculate overall clinical outcome risks including stroke and major bleeding [54,55], they did not improve prediction of stroke and major bleeding beyond the individual stroke ($CHADS_2$, CHA_2DS_2-VASc) or bleeding scores (HAS-BLED) [56]. In contrast, our study may provide insights into using a new methodology to take into account individual benefit-harm outcomes with warfarin therapy. Our PLR model calculates the specific probabilities of stroke and major bleeding at the same time, which may be more practical and acceptable in real-world clinical practice compared with using separate stroke and bleeding risk-stratification scores. Moreover, because our model produces individualized risk estimates for each patient based on various characteristics, it offers more personalized and detailed information for patients with AF rather than the population-level estimates associated with $CHADS_2$, CHA_2DS_2-VASc, and HAS-BLED scores [22]. Thus the PLR model may better facilitate patient-physician shared decision-making with regard to warfarin therapy initiation.

In our study, an unexpected inverse association between comorbid hypertension and stroke, major bleeding, and all-cause death was observed. During the model construction, we used either the ICD-9-CM codes or the antihypertensive drug surrogates including angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, thiazides, beta-blockers, calcium channel blockers, and other antihypertensive purchases, to identify hypertension comorbidity (**Supplemental Table 5**). Additionally, we ran two *post-hoc* sensitivity analyses using different methods to imply hypertension diagnosis: ICD-9-CM codes only, and both ICD-9-CM codes and antihypertensive drug purchases. These two methods yielded the same predictors included in the PLR and Cox model with extremely similar coefficients (**Supplemental Table 6**). Moreover, removing hypertension from the

model entirely also yielded similar results (**Supplemental Table 7 for the PLR model; Supplemental Table 8 for the Cox model**). Therefore, the unexpected relationship between hypertension and outcomes requires further exploration.

The strengths of our study include the use of a large sample of patients with AF to construct and validate the prediction model. Moreover, model building, assessment, and validation included rigorous and detailed statistical analyses. Another strength is the efforts in controlling bias in study design and data analyses to preclude misleading predictors from being included into the models. Nevertheless, our study also has several limitations. The majority of the data used in this study were from ICD-9-CM codes only without confirmatory chart review of the diagnosis. Thus data accuracy for baseline comorbidities may be less than optimal. Likewise, the incidence rates of stroke and major bleeding may be over- or underestimated. This could lead to false positive/negative values and weaken the findings based on the data. Additionally, we intended to predict four outcome quadrants (**Table 1**). However, the sample size for the patients with harm and no benefit was insufficient for model construction. Another limitation is lack of data from contemporary non-KPCO cohorts for model validation; thereby, potentially limiting the generalizability of the prediction model [26].

Conclusions

In this study, we introduce a new methodology for predicting individual combined benefit and harm outcomes associated with warfarin therapy for patients with AF. Should this approach be validated in other patient populations, it has potential advantages over existing risk stratification approaches as a patient-physician aid for shared decision-making.

Acknowledgement

We thank Dr. Jennifer A. Pereira for her substantial preparatory contribution in her unpublished doctoral thesis [23] to this study.

Funding statement

GL received a Father Sean O'Sullivan Research Award, the Research Institute of St. Joseph's Healthcare Hamilton, and a doctoral award from the CSC. This study received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author Contributions

Conceived and designed the study: GL, LT and AH; Collected data: TD; Analyzed data: GL and JC; Interpreted findings and drafted the manuscript: GL, LT, TD, DMW and AH; Critical revisions of the manuscript: TD, DMW and MAHL. All authors approved the final manuscript.

Competing interests

The authors declare no competing interest exits.

References

1. Rockson SG, Albers GW (2004) Comparing the guidelines: anticoagulation therapy to optimize stroke prevention in patients with atrial fibrillation. J Am Coll Cardiol 43: 929-935.

2. Camm AJ, Kirchhof P, Lip GY, Schotten U, Savelieva I, et al. (2010) Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). Eur Heart J 31: 2369-2429.

3. Hart RG, Halperin JL, Pearce LA, Anderson DC, Kronmal RA, et al. (2003) Lessons from the Stroke Prevention in Atrial Fibrillation trials. Ann Intern Med 138: 831-838.

4. Cairns JA, Connolly S, McMurtry S, Stephenson M, Talajic M (2011) Canadian Cardiovascular Society atrial fibrillation guidelines 2010: prevention of stroke and systemic thromboembolism in atrial fibrillation and flutter. Can J Cardiol 27: 74-90.

5. Wells G, Coyle D, Cameron C, Steiner S, Coyle K, et al. Safety, effectiveness, and cost-effectiveness of new oral anticoagulants compared with warfarin in preventing stroke and other cardiovascular events in patients with atrial fibrillation [Internet]. Available from: http://www.cadth.ca/media/pdf/NOAC_Therapeutic_Review_final_report.pdf. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2012 Apr 9 173 p (CADTH Therapeutic Review).

6. Holbrook A, Labiris R, Goldsmith CH, Ota K, Harb S, et al. (2007) Influence of decision aids on patient preferences for anticoagulant therapy: a randomized trial. CMAJ 176: 1583-1587.

7. Ansell J (2010) Warfarin versus new agents: interpreting the data. Hematology Am Soc Hematol Educ Program 2010: 221-228.

8. Arepally G, Bauer KA, Bhatt DL, Merli GJ, Naccarelli GV, et al. (2010) The use of antithrombotic therapies in the prevention and treatment of arterial and venous thrombosis: a survey of current knowledge and practice supporting the need for clinical education. Crit Pathw Cardiol 9: 41-48.

9. McCrory DC, Matchar DB, Samsa G, Sanders LL, Pritchett EL (1995) Physician attitudes about anticoagulation for nonvalvular atrial fibrillation in the elderly. Arch Intern Med

155: 277-281.

10. Holbrook A, Pullenayegum E, Troyan S, Nikitovic M, Crowther M (2013) Personalized Benefit-Harm Information Influences Patient Decisions Regarding Warfarin. J Popul Ther Clin Pharmacol 20: e406-e415.
11. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ (2010) Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest 137: 263-272.
12. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, et al. (2001) Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA 285: 2864-2870.
13. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, et al. (2010) A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest 138: 1093-1100.
14. Lip GY (2011) Implications of the CHA(2)DS(2)-VASc and HAS-BLED Scores for thromboprophylaxis in atrial fibrillation. Am J Med 124: 111-114.
15. Olesen JB, Lip GY, Hansen PR, Lindhardsen J, Ahlehoff O, et al. (2011) Bleeding risk in 'real world' patients with atrial fibrillation: comparison of two established bleeding prediction schemes in a nationwide cohort. J Thromb Haemost 9: 1460-1467.
16. Lip GY, Frison L, Halperin JL, Lane DA (2011) Comparative Validation of a Novel Risk Score for Predicting Bleeding Risk in Anticoagulated Patients With Atrial FibrillationThe HAS-BLED (Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile INR, Elderly, Drugs/Alcohol Concomitantly) Score. Journal of the American College of Cardiology 57: 173-180.
17. Li G, Holbrook A, Delate T, Witt DM, Levine MA, et al. (2015) Prediction of individual combined benefit and harm for patients with atrial fibrillation considering warfarin therapy: a study protocol. BMJ Open 5: e009518.
18. Singer DE, Chang Y, Fang MC, Borowsky LH, Pomernacki NK, et al. (2009) The net clinical benefit of warfarin anticoagulation in atrial fibrillation. Ann Intern Med 151: 297-305.

19. Friberg L, Rosenqvist M, Lip GY (2012) Net clinical benefit of warfarin in patients with atrial fibrillation: a report from the Swedish atrial fibrillation cohort study. Circulation 125: 2298-2307.

20. Olesen JB, Lip GY, Lindhardsen J, Lane DA, Ahlehoff O, et al. (2011) Risks of thromboembolism and bleeding with thromboprophylaxis in patients with atrial fibrillation: A net clinical benefit analysis using a 'real world' nationwide cohort study. Thromb Haemost 106: 739-749.

21. BECG CB, Gray R (1984) Calculation of polychotomous logistic regression parameters using individualized regressions. Biometrika 71: 11-18.

22. Wijesinha A, Begg CB, Funkenstein HH, McNeil BJ (1982) Methodology for the differential diagnosis of a complex data set. A case study using data from routine CT scan examinations. Medical decision making: an international journal of the Society for Medical Decision Making 3: 133-154.

23. Pereira JA (2008 (unpublished doctoral dissertation)) Methods to predict individualized combined benefit/harm patient profiles for warfarin [Unpublished doctoral thesis]. Toronto, Canada: University of Toronto.

24. Witt DM, Sadler MA, Shanahan RL, Mazzoli G, Tillman DJ (2005) Effect of a centralized clinical pharmacy anticoagulation service on the outcomes of anticoagulation therapy. Chest 127: 1515-1522.

25. Zhou Z, Rahme E, Abrahamowicz M, Pilote L (2005) Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods. American journal of epidemiology 162: 1016-1023.

26. Altman DG, Vergouwe Y, Royston P, Moons KG (2009) Prognosis and prognostic research: validating a prognostic model. BMJ: British Medical Journal 338: 1432-1435.

27. Suissa S (2008) Immortal time bias in pharmacoepidemiology. American journal of epidemiology 167: 492-499.

28. Lévesque LE, Hanley JA, Kezouh A, Suissa S (2010) Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. Bmj 340: 907-911.

29. Witt DM, Delate T, Clark NP, Martell C, Tran T, et al. (2009) Outcomes and predictors of very stable INR control during chronic anticoagulation therapy. Blood 114: 952-956.

30. Schulman S, Kearon C (2005) Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost 3: 692-694.

31. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40: 373-383.

32. Holbrook AM, Pereira JA, Labiris R, McDonald H, Douketis JD, et al. (2005) Systematic overview of warfarin and its drug and food interactions. Arch Intern Med 165: 1095-1106.

33. Holbrook A, Schulman S, Witt DM, Vandvik PO, Fish J, et al. (2012) Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 141: e152S.

34. Fisher LD, Lin DY (1999) Time-dependent covariates in the Cox proportional-hazards regression model. Annual review of public health 20: 145-157.

35. Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association 94: 496-509.

36. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, et al. (2012) Risk prediction models: II. External validation, model updating, and impact assessment. Heart 98: 691-698.

37. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y (2008) Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 61: 76-86.

38. Ho WH, Lee KT, Chen HY, Ho TW, Chiu HC (2012) Disease-free survival after hepatic resection in hepatocellular carcinoma patients: a prediction approach using artificial neural network. PLoS One 7: e29179.

39. Bilen K, Ponzer S, Ottosson C, Castren M, Pettersson H (2013) Deliberate self-harm patients in the emergency department: who will repeat and who will not? Validation

and development of clinical decision rules. Emerg Med J 30: 650-656.

40. Brasier AR, Ju H, Garcia J, Spratt HM, Victor SS, et al. (2012) A three-component biomarker panel for prediction of dengue hemorrhagic fever. Am J Trop Med Hyg 86: 341-348.
41. Ferreira D, Oliveira A, Freitas A (2012) Applying data mining techniques to improve diagnosis in neonatal jaundice. BMC Med Inform Decis Mak 12: 143.
42. Cheong JY, Um SH, Seo YS, Shin SS, Park RW, et al. (2012) A practical scoring system for predicting cirrhosis in patients with chronic viral hepatitis. Hepatogastroenterology 59: 2592-2597.
43. Minne L, Eslami S, de Keizer N, de Jonge E, de Rooij SE, et al. (2012) Statistical process control for monitoring standardized mortality ratios of a classification tree model. Methods Inf Med 51: 353-358.
44. Hart RG, Pearce LA, Aguilar MI (2007) Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med 146: 857-867.
45. Wysowski DK, Nourjah P, Swartz L (2007) Bleeding complications with warfarin use: a prevalent adverse effect resulting in regulatory action. Arch Intern Med 167: 1414-1419.
46. Morgan CL, McEwan P, Tukiendorf A, Robinson PA, Clemens A, et al. (2009) Warfarin treatment in patients with atrial fibrillation: observing outcomes associated with varying levels of INR control. Thromb Res 124: 37-41.
47. Olson KL, Wood MD, Delate T, Lash LJ, Rasmussen J, et al. (2014) Positive predictive values of ICD-9 codes to identify patients with stroke or TIA. Am J Manag Care 20: e27-34.
48. Cunningham A, Stein CM, Chung CP, Daugherty JR, Smalley WE, et al. (2011) An automated database case definition for serious bleeding related to oral anticoagulant use. Pharmacoepidemiol Drug Saf 20: 560-566.
49. Ince B, Benbir G, Gozubatik-Celik G (2014) Should HAS-BLED scoring be revised for better risk estimation in patients with intracerebral hemorrhage? Expert Rev Cardiovasc Ther 12: 929-931.

50. Poli D, Antonucci E, Testa S, Cosmi B, Palareti G, et al. (2013) The predictive ability of bleeding risk stratification models in very old patients on vitamin K antagonist treatment for venous thromboembolism: results of the prospective collaborative EPICA study. J Thromb Haemost 11: 1053-1058.

51. Chao TF, Liu CJ, Wang KL, Lin YJ, Chang SL, et al. (2015) Should atrial fibrillation patients with 1 additional risk factor of the CHA2DS2-VASc score (beyond sex) receive oral anticoagulation? J Am Coll Cardiol 65: 635-642.

52. Donze J, Rodondi N, Waeber G, Monney P, Cornuz J, et al. (2012) Scores to predict major bleeding risk during oral anticoagulation therapy: a prospective validation study. Am J Med 125: 1095-1102.

53. Olesen JB, Lip GY, Hansen ML, Hansen PR, Tolstrup JS, et al. (2011) Validation of risk stratification schemes for predicting stroke and thromboembolism in patients with atrial fibrillation: nationwide cohort study. BMJ 342: d124.

54. Lip GY, Lane DA, Buller H, Apostolakis S (2013) Development of a novel composite stroke and bleeding risk score in patients with atrial fibrillation: the AMADEUS Study. Chest 144: 1839-1847.

55. Banerjee A, Fauchier L, Bernard-Brunet A, Clementy N, Lip GY (2014) Composite risk scores and composite endpoints in the risk prediction of outcomes in anticoagulated patients with atrial fibrillation. The Loire Valley Atrial Fibrillation Project. Thromb Haemost 111: 549-556.

56. Dzeshka MS, Lane DA, Lip GY (2014) Stroke and bleeding risk in atrial fibrillation: navigating the alphabet soup of risk-score acronyms (CHADS2 , CHA2 DS2 -VASc, R2 CHADS2 , HAS-BLED, ATRIA, and more). Clin Cardiol 37: 634-644.

Table and Figure legends:

Table 1. Warfarin's combined benefit and harm outcomes

Table 2. Characteristics of study patients stratified by warfarin users versus non-users in derivation and validation cohort

Table 3. Outcomes until study outcome end date between warfarin users and non-users in KPCO-I and KPCO-II cohorts

Table 4. Results of the original PLR model and bootstrap analyses for stroke and major bleeding in KPCO-I cohort

Table 5. Results of the Cox model for death in the KPCO-I cohort

Table 6. Sensitivity analysis results from competing risk analysis for stroke and bleeding based on survival analysis in KPCO-I cohort

Table 7. Model performance of PLR model for stroke and major bleeding and Cox model for death in KPCO-I and KPCO-II cohorts

Supplemental Figure 1. Flow diagram of selecting patients for analyses

Supplemental Table 1. Characteristics of study patients stratified by taking versus not taking warfarin for the whole cohort

Supplemental Figure 2. Kaplan-Meier survival curves for death in the derivation and validation cohorts

Supplemental Table 2. Rates of stroke and major bleeding in the KPCO cohorts stratified by CHA_2DS_2-VASc and HAS-BLED scores

Supplemental Table 3. Sensitivity analysis results from multivariable model to assess time-varying effect of warfarin on stroke, major bleeding and death

Supplemental Figure 3. Calibration curve in the PLR model for stroke in the derivation cohort

Supplemental Figure 4. Calibration curve in the PLR model for major bleeding in the derivation cohort

Supplemental Figure 5. Calibration curve in the Cox model for death in the derivation cohort

Supplemental Table 4. Updates of the models' intercepts and the regression coefficients for external validation in the KPCO-Ⅱ cohort

Supplemental Figure 6. Calibration curve in the PLR model for stroke in the validation cohort

Supplemental Figure 7. Calibration curve in the PLR model for major bleeding in the validation cohort

Supplemental Figure 8. Calibration curve in the Cox model for death in the validation cohort

Supplemental Table 5. Hypertensive drugs as surrogates for hypertension diagnosis

Supplemental Table 6. Results for effect of hypertension in the PLR and Cox model using

different data on hypertension in the KPCO-I cohort

Supplemental Table 7. Sensitivity analysis leaving hypertension out of the PLR model for stroke and major bleeding in the KPCO-I cohort

Supplemental Table 8. Sensitivity analysis leaving hypertension out of the Cox model for death in the KPCO-I cohort

Table 1. Warfarin's combined benefit and harm outcomes

	Harm **(major bleeding)**	**No harm** **(no major bleeding)**
Benefit **(no stroke)**	No stroke/major bleeding	No stroke/no major bleeding
No benefit **(stroke)**	Stroke/major bleeding	Stroke/no major bleeding

Table 2. Characteristics of study patients stratified by warfarin users versus non-users in derivation and validation cohort

Baseline Characteristics	Total participants (n=9074)	**KPCO-I (n=4632)[1]**			**KPCO-II (n=4442)[2]**		
		Warfarin users (n=2316)	Warfarin non-users (n=2316)	P-value	Warfarin users (n=2221)	Warfarin non-users (n=2221)	P-value
Age: mean (SD), years	71.7 (13.00)	72.3 (10.74)	70.5 (15.26)	<0.001	72.9 (10.64)	71.3 (14.54)	<0.001
Female: n (%)	4199 (46.28)	1229 (53.07)	1209 (52.20)	0.556	1275 (57.41)	1162 (52.32)	<0.001
Comorbidities: n (%)							
Congestive heart failure	1064 (11.73)	286 (12.35)	220 (9.50)	0.002	325 (14.63)	233 (10.49)	<0.001
Hypertension	7132 (78.60)	2024 (87.39)	1609 (69.47)	<0.001	1957 (88.11)	1542 (69.43)	<0.001
Diabetes	1759 (19.39)	428 (18.48)	391 (16.88)	0.154	509 (22.92)	431 (19.41)	0.004
Prior stroke/TIA	539 (5.94)	122 (5.27)	112 (4.84)	0.502	197 (8.87)	108 (4.86)	<0.001
Myocardial infarction	516 (5.69)	93 (4.02)	94 (4.06)	0.941	183 (8.24)	146 (6.57)	0.034
Peripheral vascular disease	615 (6.78)	138 (5.96)	129 (5.57)	0.571	183 (8.24)	165 (7.43)	0.315
Renal disease	1146 (12.63)	174 (7.51)	219 (9.46)	0.018	406 (18.28)	347 (15.62)	0.018
Liver disease	20 (0.22)	3 (0.13)	4 (0.17)	0.705#	2 (0.09)	11 (0.50)	0.022#
Prior major bleeding	260 (2.87)	74 (3.20)	103 (4.45)	0.026	42 (1.89)	41 (1.85)	0.912

Anemia	657 (7.24)	142 (6.13)	189 (8.16)	0.007	127 (5.72)	199 (8.96)	<0.001
Alcohol abuse	119 (1.31)	15 (0.65)	34 (1.47)	0.006	24 (1.08)	46 (2.07)	0.008
Other cerebrovascular disease	194 (2.14)	38 (1.64)	43 (1.86)	0.575	58 (2.61)	55 (2.48)	0.775
Dementia	21 (0.23)	2 (0.09)	4 (0.17)	0.687#	1 (0.05)	14 (0.63)	<0.001#
Chronic pulmonary disease	468 (5.16)	115 (4.97)	89 (3.84)	0.063	153 (6.89)	111 (5.00)	0.008
Rheumatic disease	245 (2.70)	67 (2.89)	56 (2.42)	0.315	58 (2.61)	64 (2.88)	0.582
Peptic ulcer disease	57 (0.63)	12 (0.52)	15 (0.65)	0.563	10 (0.45)	20 (0.90)	0.067
Hemiplegia or paraplegia	33 (0.36)	5 (0.22)	9 (0.39)	0.423	7 (0.32)	12 (0.54)	0.250
Malignancy[3]	816 (8.99)	196 (8.46)	220 (9.50)	0.217	170 (7.65)	230 (10.36)	0.002
AIDS or HIV	0	0	0	-	0	0	-
CHA_2DS_2–VASc score	2.99 (1.56)	3.09 (1.43)	2.73 (1.64)	<0.001	3.29 (1.51)	2.85 (1.61)	<0.001
HAS-BLED score[4]	1.73 (0.88)	1.80 (0.73)	1.54 (0.94)	<0.001	1.96 (0.82)	1.63 (0.95)	<0.001
Concurrent medication use interacting with warfarin: n (%)							
Other anticoagulants	123 (1.36)	34 (1.47)	38 (1.64)	0.635	25 (1.13)	26 (1.17)	0.888
Antiplatelets	836 (9.21)	184 (7.94)	248 (10.71)	0.001	194 (8.73)	210 (9.46)	0.404
NSAIDs	766 (8.44)	253 (10.92)	197 (8.51)	0.006	183 (8.24)	133 (5.99)	0.004

Antibiotics	1726 (19.02)	496 (21.42)	432 (18.65)	0.019	471 (21.21)	327 (14.72)	<0.001
Antifungals	169 (1.86)	38 (1.64)	48 (2.07)	0.276	34 (1.53)	49 (2.21)	0.097
Antitubercular agents	1 (0.01)	1 (0.04)	0 (0)	1.000#	0	0	-
Cardiac drugs	1706 (18.80)	571 (24.65)	323 (13.95)	<0.001	559 (25.17)	253 (11.39)	<0.001
Antilipemic drugs	81 (0.89)	16 (0.69)	13 (0.56)	0.576	31 (1.40)	21 (0.95)	0.163
Antidepressants	1059 (11.67)	263 (11.36)	276 (11.92)	0.551	284 (12.79)	236 (10.63)	0.025
Other CNS drugs	52 (0.57)	13 (0.56)	15 (0.65)	0.705	15 (0.68)	9 (0.41)	0.219
GI drugs	1836 (20.23)	499 (21.55)	403 (17.40)	<0.001	538 (24.22)	396 (17.83)	<0.001
Other drug[5]	255 (2.81)	58 (2.50)	28 (1.21)	0.001	111 (5.00)	58 (2.61)	<0.001
Laboratory information: mean (SD)							
Serum creatinine, mg/dl	1.18 (0.78)	1.18 (0.81)	1.24 (0.88)	0.077	1.14 (0.57)	1.17 (0.82)	0.232
INR	1.49 (0.75)	1.60 (0.85)	1.36 (0.67)	<0.001	1.62 (0.79)	1.28 (0.53)	<0.001
Albumin, g/dl	3.85 (0.70)	3.91 (0.65)	3.89 (0.65)	0.553	3.82 (0.71)	3.79 (0.78)	0.300
Hemoglobin, g/dl	13.74	14.00 (2.14)	13.86 (2.14)	0.065	13.52 (2.30)	13.60 (2.22)	0.345

	(2.21)						

SD = standard deviation; TIA = transient ischemic attack; AIDS or HIV = acquired immune deficiency syndrome or human immunodeficiency virus infection; NSAIDs = non-steroidal anti-inflammatory drugs; CNS drugs = central nervous system drugs; INR = international normalized ratio.

[1]Median follow-up: 652 days (interquartile range: 299 to 1068); [2]Median follow-up: 628 days (interquartile range: 293 to 1036);

[3]Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin;

[4] No data on labile INR to calculate the HAS-BLED score; [5] Other drug included tramadol

Fisher's exact test

Table 3. Outcomes until study outcome end date between warfarin users and non-users in KPCO-I and KPCO-II cohorts

Outcomes	Total participants (n=9074)	**KPCO-I (n=4632)**			**KPCO-II (n=4442)**		
		Warfarin users (n=2316)	Warfarin non-users (n=2316)	P-value	Warfarin users (n=2221)	Warfarin non-users (n=2221)	P-value
Stroke, n (%)	278 (3.06)	65 (2.81)	71 (3.07)	0.602	71 (3.20)	71 (3.20)	1.000
Major bleeding, n (%)	453 (4.99)	181 (7.82)	99 (4.27)	<0.001	106 (4.77)	67 (3.02)	0.003
Death, n (%)	2186 (24.09)	442 (19.08)	752 (32.47)	<0.001	355 (15.98)	637 (28.68)	<0.001

Table 4. Results of the original PLR model and bootstrap analyses for stroke and major bleeding in KPCO-I cohort

Predictors	Stroke vs. neither event (OR with 95%CI, p-value)		Major bleeding vs. neither event (OR with 95%CI, p-value)	
	Original model	Bootstrap model	Original model	Bootstrap model
Intercept: coefficient β , p-value	-3.76, <0.001	-3.91, <0.001	-4.21, < 0.001	-4.09, <0.001
Age[1]: years	1.02 (1.00-1.04), 0.013	1.02 (1.01-1.04), 0.010	1.02 (1.01-1.03), 0.001	1.02 (1.01-1.03), 0.001
Female	1.51 (1.06-2.13), 0.025	1.53 (1.05-2.22), 0.024	0.73 (0.56-0.94), 0.015	0.74 (0.56-0.97), 0.028
Warfarin	0.94 (0.66-1.34),0.711	0.97 (0.69-1.43), 0.789	1.71 (1.32-2.22), < 0.001	1.89 (1.46-2.44), <0.001
Other cerebrovascular disease	4.76 (2.42-9.37), <0.001	4.85 (2.34-10.03), <0.001	1.36 (0.60-3.15), 0.469	1.33 (0.39-4.47), 0.338
Congestive heart failure	1.25 (0.71-2.22), 0.434	1.30 (0.68-2.43), 0.427	1.59 (1.14-2.23), 0.007	1.59 (1.14-2.24), 0.008
Hypertension	0.88 (0.56-1.39), 0.587	0.81 (0.54-1.18), 0.326	0.94 (0.66-1.33), 0.712	0.92 (0.64-1.31), 0.695
Diabetes	1.13 (0.72-1.79), 0.598	1.15 (0.70-1.90),	1.21 (0.89-1.65), 0.233	1.25 (0.89-1.74),

		0.412		0.177
Prior major bleeding	1.12 (0.50-2.52), 0.782	1.06 (0.42-2.70), 0.783	1.49 (0.87-2.54), 0.147	1.48 (0.84-2.62), 0.149
Prior stroke	2.04 (1.16-3.57), 0.013	2.08 (1.14-3.78), 0.014	0.73 (0.40-1.34), 0.313	0.70 (0.38-1.31), 0.274
Renal disease	1.35 (0.74-2.45), 0.329	1.40 (0.75-2.61), 0.273	1.51 (1.01-2.30), 0.046	1.51 (0.98-2.32), 0.050
Concurrent use of antibiotics	0.98 (0.64-1.51), 0.916	0.97 (0.62-1.53), 0.803	1.81 (1.39-2.37), < 0.001	1.81 (1.38-2.40), <0.001
Concurrent use of antiplatelets	1.71 (1.05-2.77), 0.030	1.67 (1.01-2.76), 0.045	1.57 (1.09-2.27), 0.017	1.57 (1.06-2.33), 0.018
Concurrent use of gastrointestinal medications	1.19 (0.75-1.88), 0.459	1.18 (0.74-1.89), 0.398	1.77 (1.35-2.33), < 0.001	1.79 (1.34-2.39), <0.001

PLR = polytomous logistic regression; OR= odds ratio; CI = confidence interval

[1] Used as per one-year change

Table 5. Results of the Cox model for death in the KPCO-I cohort

Predictors	All-cause death (n=1194)			
	Original model		Bootstrap model	
	HR (95% CI)	P-value	HR (95% CI)	P-value
Age[1]: years	1.06 (1.06-1.07)	<0.001	1.06 (1.06-1.07)	<0.001
Warfarin	0.55 (0.49-0.62)	< 0.001	0.52 (0.47-0.59)	<0.001
Anemia	1.91 (1.61-2.26)	< 0.001	1.91 (1.58-2.30)	<0.001
Other cerebrovascular disease	1.69 (1.22-2.35)	0.001	1.73 (1.22-2.48)	0.002
Congestive heart failure	1.50 (1.29-1.75)	< 0.001	1.49 (1.28-1.76)	<0.001
Diabetes	1.49 (1.29-1.70)	< 0.001	1.53 (1.32-1.76)	<0.001
Hypertension	0.76 (0.66-0.85)	< 0.001	0.76 (0.67-0.86)	<0.001
Prior major bleeding	1.37 (1.07-1.76)	0.012	1.39 (1.09-1.76)	0.007
Malignancy[2]	1.87 (1.59-2.19)	<0.001	1.86 (1.57-2.22)	<0.001
Concurrent use of antifungals	1.56 (1.11-2.17)	0.009	1.56 (1.14-2.13)	0.006
Concurrent use of antidepressants	1.22 (1.04-1.45)	0.013	1.19 (1.03-1.39)	0.015

HR = hazard ratio; CI = confidence interval

[1] Used as per one-year change; [2] Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin

Table 6. Sensitivity analysis results from competing risk analysis for stroke and bleeding based on survival analysis in KPCO-I cohort

Predictors	All-cause death as a competing risk[1]	
	Stroke (n=136) vs. no stroke (SHR with 95% CI, p-value)	Major bleeding (n=280) vs. no major bleeding (SHR with 95% CI, p-value)
Age[2]: years	1.01 (1.00-1.03), 0.043	1.01 (1.00-1.03), 0.028
Female	1.56 (1.11-2.22), 0.012	0.75 (0.59-0.96), 0.023
Warfarin	0.94 (0.66-1.36), 0.759	1.84 (1.43-2.36), <0.001
Other cerebrovascular disease	4.31 (2.27-8.20), <0.001	1.21 (0.57-2.58), 0.624
Congestive heart failure	1.35 (0.76-2.38), 0.296	1.52 (1.10-2.11), 0.011
Hypertension	0.92 (0.58-1.44), 0.705	0.93 (0.68-1.29), 0.674
Diabetes	1.07 (0.68-1.69), 0.763	1.18 (0.88-1.59), 0.261
Prior major bleeding	1.06 (0.50-2.26), 0.885	1.40 (0.87-2.27), 0.166
Prior stroke	1.97 (1.14-3.42), 0.015	0.71 (0.39-1.29), 0.259
Renal disease	1.20 (0.67-2.16), 0.546	1.37 (0.93-2.01), 0.108
Concurrent use of antibiotics	0.90 (0.59-1.38), 0.629	1.70 (1.31-2.21), <0.001
Concurrent use of antiplatelets	1.65 (1.02-2.68), 0.042	1.47 (1.02-2.11), 0.037
Concurrent use of gastrointestinal medications	1.25 (0.79-1.95), 0.340	1.75 (1.34-2.28), <0.001

SHR = subdistribution hazard ratio; CI = confidence interval

[1] The Fine and Gray proportional subdistribution hazards model was used; [2] Used as per one-year change

Table 7. Model performance of PLR model for stroke and major bleeding and Cox model for death in KPCO-I and KPCO-II cohorts

Model performance	KPCO-I (n=4632)			KPCO-II (n=4442)		
	PLR model[3]		Cox model	PLR model		Cox model
	Stroke vs. neither event	Major bleeding vs. neither event	Death vs. survival	Stroke vs. neither event	Major bleeding vs. neither event	Death vs. survival
Goodness-of-fit test statistics (p-value)[1]	8.61 (0.377)	11.08 (0.197)	14.34 (0.114)	10.32 (0.243)	7.30 (0.505)	15.01 (0.093)
Discrimination (95% CI)[2]	0.71 (0.65-0.75)	0.72 (0.68-0.75)	0.75 (0.73-0.76)	0.65 (0.60-0.69)	0.66 (0.62-0.70)	0.76 (0.74-0.77)

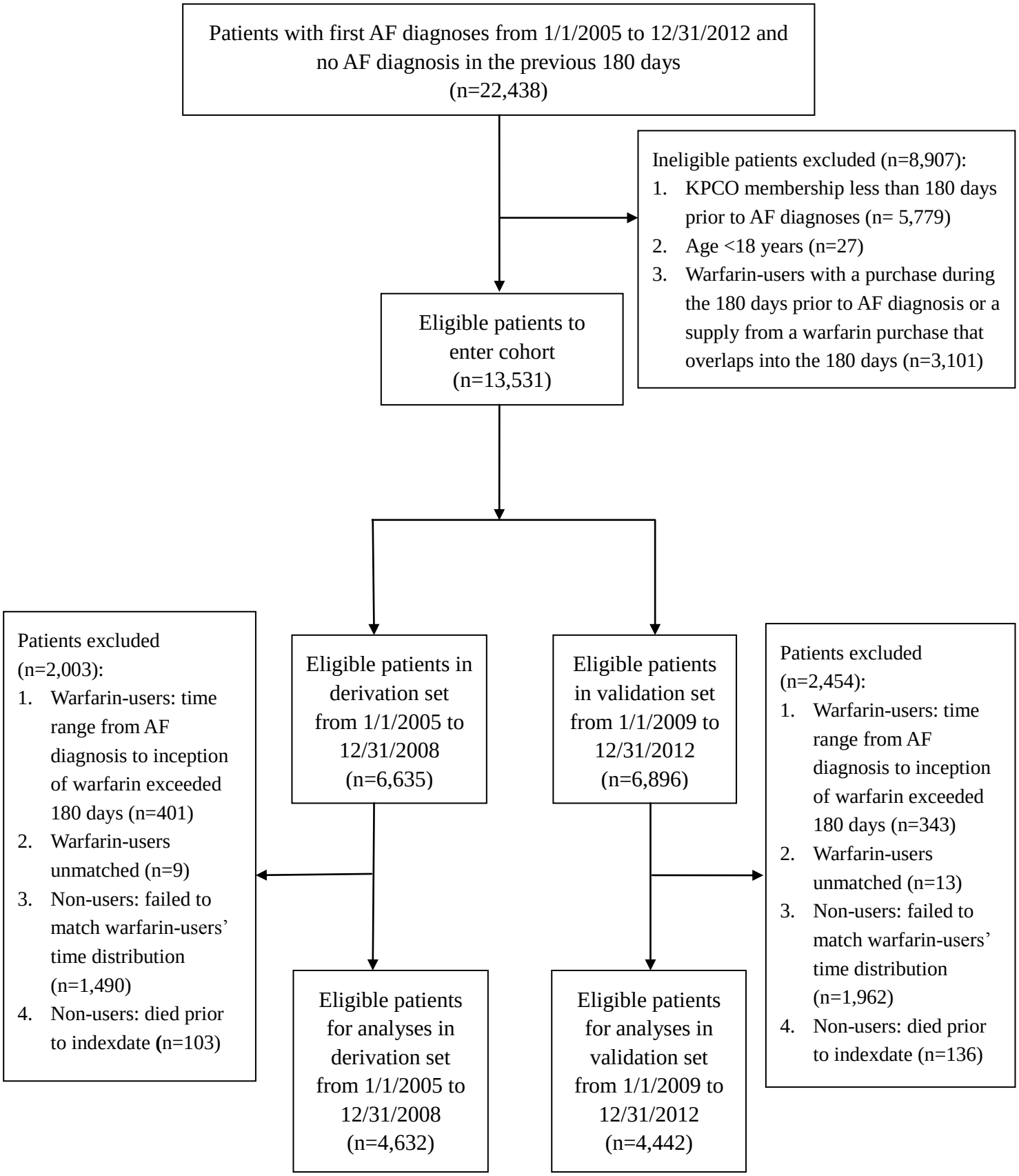
PLR = polytomous logistic regression

[1] Hosmer-Lemeshow test used for the PLR model, Groennesby and Borgan test used for the Cox model;

[2] Area under the receiver operating characteristic curves (AUC) used for the PLR model, Harrell's C index used for the Cox model;

[3] AUC from 10-fold cross-validation for stroke vs. neither event: 0.69 (0.66-0.71), for major bleeding vs. neither event: 0.71 (0.69-0.72)

Supplemental Figure 1. Flow diagram of selecting patients for analyses



Supplemental Table 1. Characteristics of study patients stratified by taking versus not taking warfarin for the whole cohort

Baseline Characteristics	**KPCO cohort**			
	Total participants (n=9074)	Warfarin-users (n=4537)	Warfarin non-users (n=4537)	P-value
Age: mean (SD), years	71.7 (13.00)	72.6 (10.69)	70.9 (14.92)	<0.001
Female: n (%)	4199 (46.28)	2033 (44.81)	2166 (47.74)	0.005
Comorbidities: n (%)				
Congestive heart failure	1064 (11.73)	611 (13.47)	453 (9.98)	<0.001
Hypertension	7132 (78.60)	3981 (87.75)	3151 (69.45)	<0.001
Diabetes	1759 (19.39)	937 (20.65)	822 (18.12)	0.002
Prior stroke/TIA	539 (5.94)	319 (7.03)	220 (4.85)	<0.001
Myocardial infarction	516 (5.69)	276 (6.08)	240 (5.29)	0.103
Peripheral vascular disease	615 (6.78)	321 (7.08)	294 (6.48)	0.260
Renal disease	1146 (12.63)	580 (12.78)	566 (12.48)	0.658
Liver disease	20 (0.22)	5 (0.11)	15 (0.33)	0.041#
Prior major bleeding	260 (2.87)	116 (2.56)	144 (3.17)	0.078
Anemia	657 (7.24)	269 (5.93)	388 (8.55)	<0.001
Alcohol abuse	119 (1.31)	39 (0.86)	80 (1.76)	<0.001
Other cerebrovascular disease	194 (2.14)	96 (2.12)	98 (2.16)	0.885
Dementia	21 (0.23)	3 (0.07)	18 (0.40)	0.002#
Chronic pulmonary disease	468 (5.16)	268 (5.91)	200 (4.41)	0.001
Rheumatic disease	245 (2.70)	125 (2.76)	120 (2.64)	0.746
Peptic ulcer disease	57 (0.63)	22 (0.48)	35 (0.77)	0.084
Hemiplegia or paraplegia	33 (0.36)	12 (0.26)	21 (0.46)	0.117
Malignancy[1]	816 (8.99)	366 (8.07)	450 (9.92)	0.002
AIDS or HIV	0	0	0	-
CHA_2DS_2–VASc score	2.99 (1.56)	3.19 (1.47)	2.79 (1.63)	<0.001

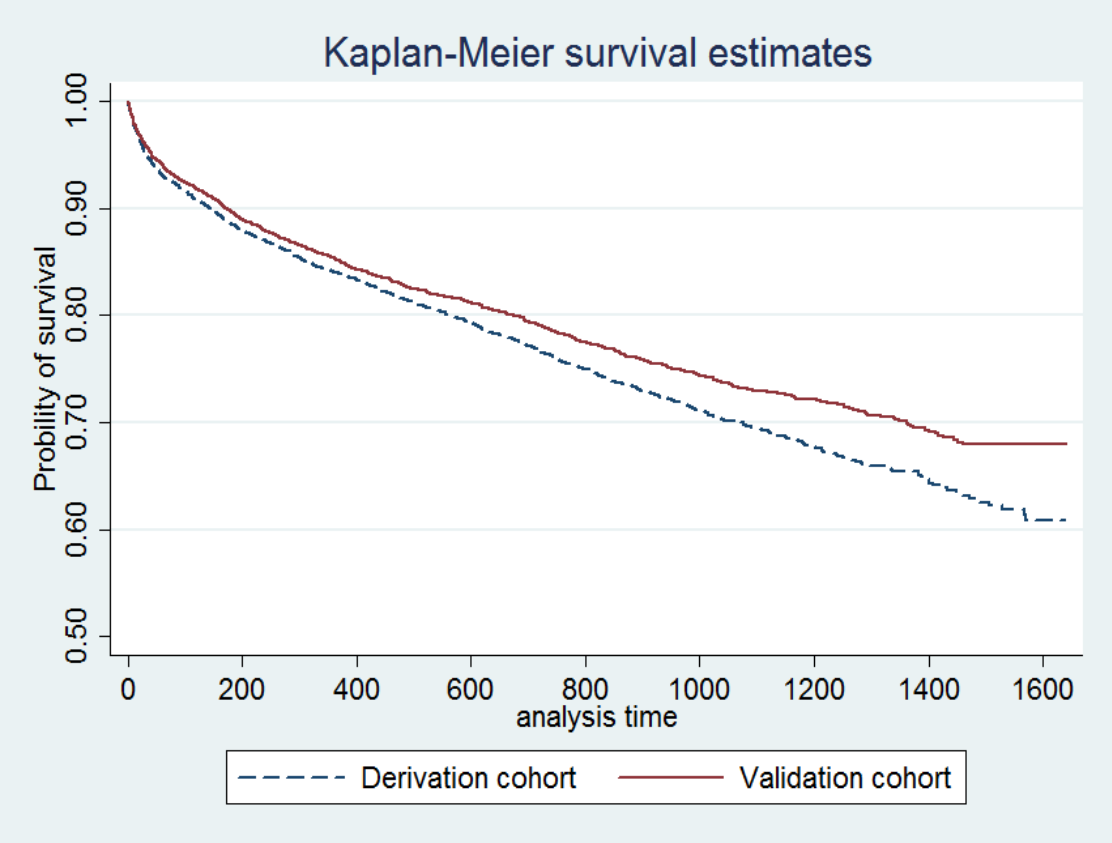
HAS-BLED score[2]	1.73 (0.88)	1.88 (0.77)	1.58 (0.95)	<0.001
Concurrent medication use interacting with warfarin: n (%)				
Other anticoagulants	123 (1.36)	59 (1.30)	64 (1.41)	0.650
Antiplatelets	836 (9.21)	378 (8.33)	458 (10.09)	0.004
NSAIDs	766 (8.44)	436 (9.61)	330 (7.27)	<0.001
Antibiotics	1726 (19.02)	967 (21.31)	759 (16.73)	<0.001
Antifungals	169 (1.86)	72 (1.59)	97 (2.14)	0.052
Antitubercular agents	1 (0.01)	1 (0.02)	0 (0)	1.000#
Cardiac drugs	1706 (18.80)	1130 (24.91)	576 (12.70)	<0.001
Antilipemic drugs	81 (0.89)	47 (1.04)	34 (0.75)	0.147
Antidepressants	1059 (11.67)	547 (12.06)	512 (11.28)	0.253
Other CNS drugs	52 (0.57)	28 (0.62)	24 (0.53)	0.578
GI drugs	1836 (20.23)	1037 (22.86)	799 (17.61)	<0.001
Other drug	255 (2.81)	169 (3.72)	86 (1.90)	<0.001
Laboratory information: mean (SD)				
Serum creatinine, mg/dl	1.18 (0.78)	1.16 (0.69)	1.20 (0.85)	0.028
INR	1.49 (0.75)	1.61 (0.82)	1.32 (0.60)	<0.001
Albumin, g/dl	3.85 (0.70)	3.86 (0.68)	3.84 (0.72)	0.288
Hemoglobin, g/dl	13.74 (2.21)	13.75 (2.24)	13.73 (2.18)	0.705

[1] Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin;

[2] No data on labile INR to calculate the HAS-BLED score;

Fisher's exact test

Supplemental Figure 2. Kaplan-Meier survival curves for death in the derivation and validation cohorts



Supplemental Table 2. Rates of stroke and major bleeding in the KPCO cohorts stratified by CHA_2DS_2-VASc and HAS-BLED scores*

	KPCO-I	KPCO-II	P-value
Stroke stratified by CHA_2DS_2–VASc score: event number/total number (%)			
0	1/258 (0.39)	1/216 (0.46)	0.998#
1	16/683 (2.34)	12/562 (2.14)	0.806
2	17/894 (1.90)	19/822 (2.31)	0.554
3	26/1078 (2.41)	29/1050 (2.76)	0.611
4	46/1076 (4.28)	42/1040 (4.04)	0.785
5	22/443 (4.97)	17/494 (3.44)	0.243
6	7/146 (4.79)	12/184 (6.52)	0.504
7	0/43 (0)	8/57 (14.04)	0.011#
8	0/9 (0)	2/17 (11.76)	0.529#
9	1/2 (50.00)	-	-
Total	136/4632 (2.94)[1]	142/4442 (3.20)[2]	0.471
Major bleeding stratified by HAS-BLED score[3]: n (%)			
0	15/427 (3.51)	6/357 (1.68)	0.114
1	61/1325 (4.60)	28/1178 (2.38)	0.003
2	157/2287 (6.86)	91/2045 (4.45)	0.001
3	39/528 (7.39)	42/759 (5.53)	0.178
4	8/62 (12.90)	6/98 (6.12)	0.139
5	0/3 (0)	0/5 (0)	-
6	-	-	-
7	-	-	-
8	-	-	-
Total	280/4632 (6.04)[4]	173/4442 (3.89)[5]	<0.001

* Patients' CHA_2DS_2–VASc score (minimum to maximum): 0 to 9; HAS-BLED score (minimum to maximum): 0 to 5

[1]P-value for trend < 0.001; [2]P-value for trend < 0.001; [3]No data on labile INR to calculate the HAS-BLED score;

[4]P-value for trend < 0.001; [5]P-value for trend < 0.001; # Fisher's exact test

Supplemental Table 3. Sensitivity analysis results from multivariable model to assess time-varying effect[1] of warfarin on stroke, major bleeding and death

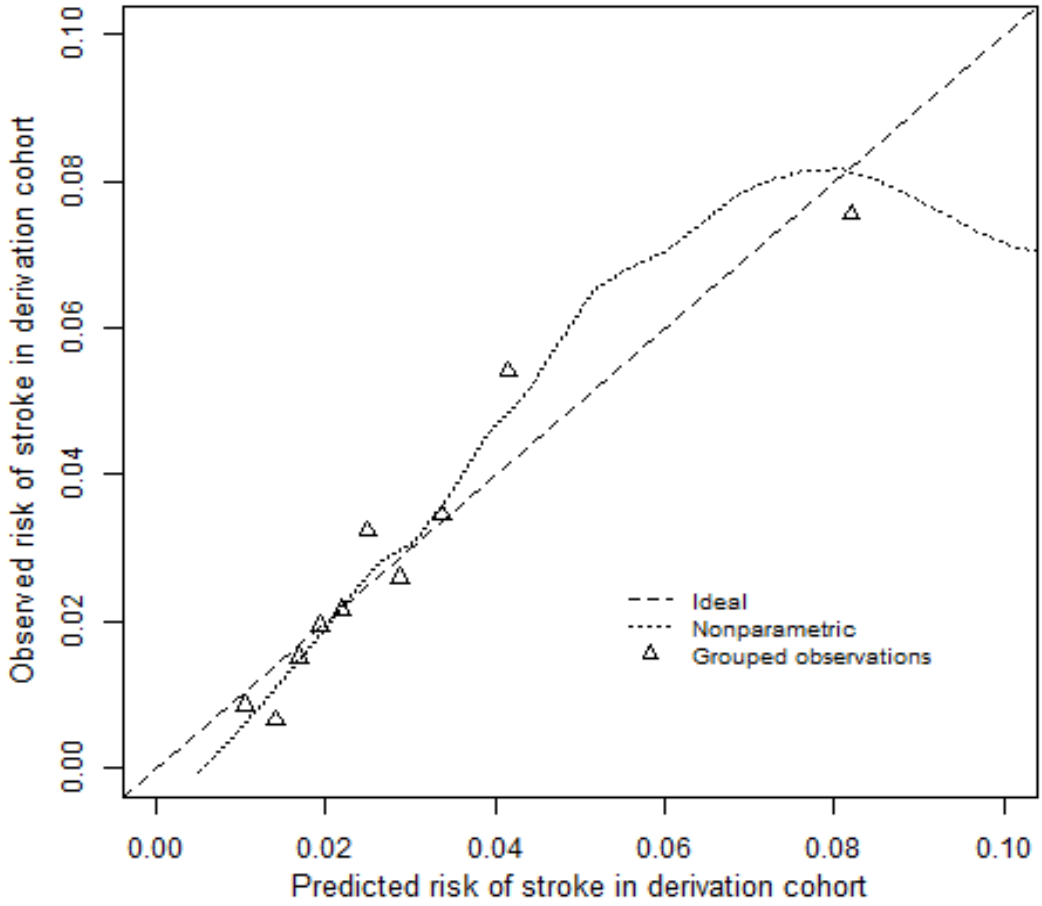
Predictor	Stroke[2] (HR with 95% CI, p-value)	Major bleeding[2] (HR with 95% CI, p-value)	Death[3] (HR with 95% CI, p-value)
Derivation cohort (n=4632)			
Warfarin	0.87 (0.57-1.32), 0.442	1.70 (1.28-2.26), <0.001	0.36 (0.28-0.45), <0.001
Validation cohort (n=4442)			
Warfarin	0.89 (0.57-1.40), 0.618	1.24 (0.78-1.96), 0.363	0.47 (0.37-0.58), <0.001

[1] Three Cox models were used to assess the time-dependent effect of warfarin on stroke, major bleeding and death, respectively;

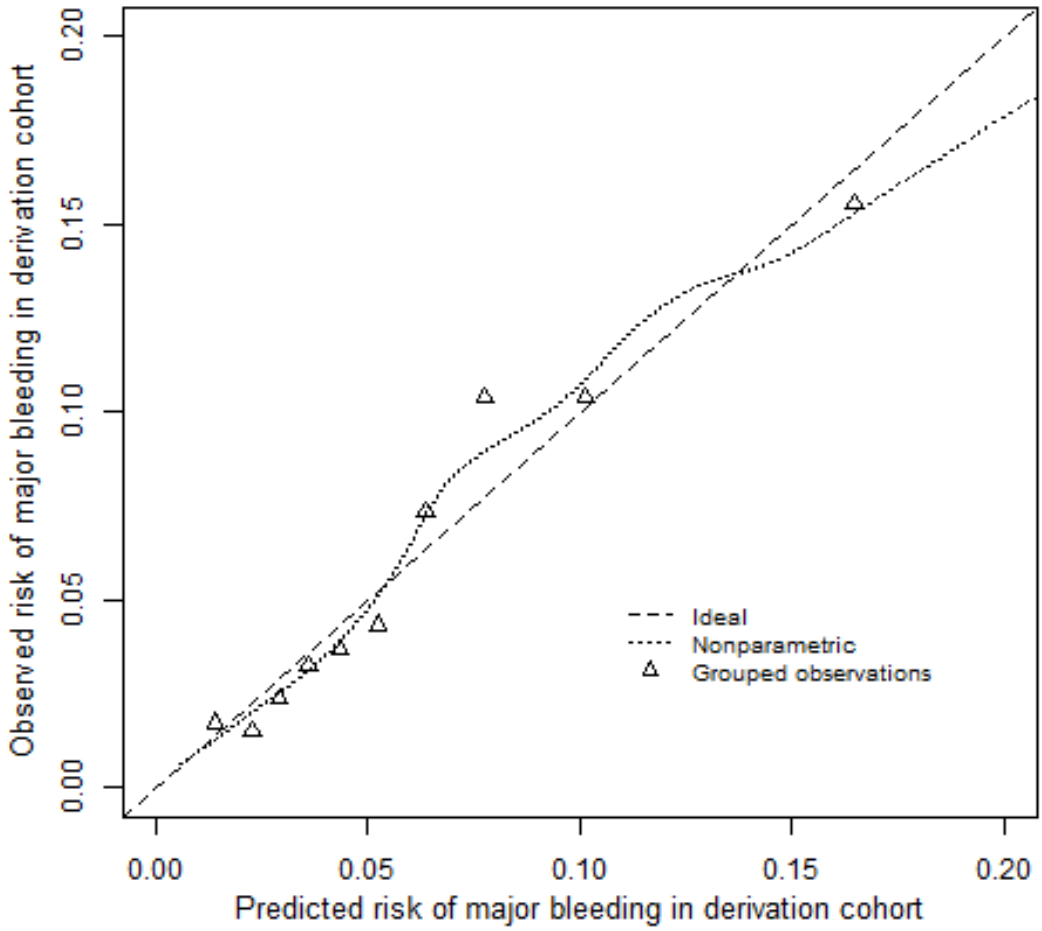
[2] Adjusted for age, sex, other cerebrovascular disease, congestive heart failure, hypertension, diabetes, prior major bleeding, prior stroke, renal disease, concurrent use of antibiotics, concurrent use of antiplatelets, and concurrent use of gastrointestinal medications;

[3] Adjusted for age, anemia, other cerebrovascular disease, congestive heart failure, hypertension, diabetes, prior major bleeding, malignancy, concurrent use of antifungals, and concurrent use of antidepressants

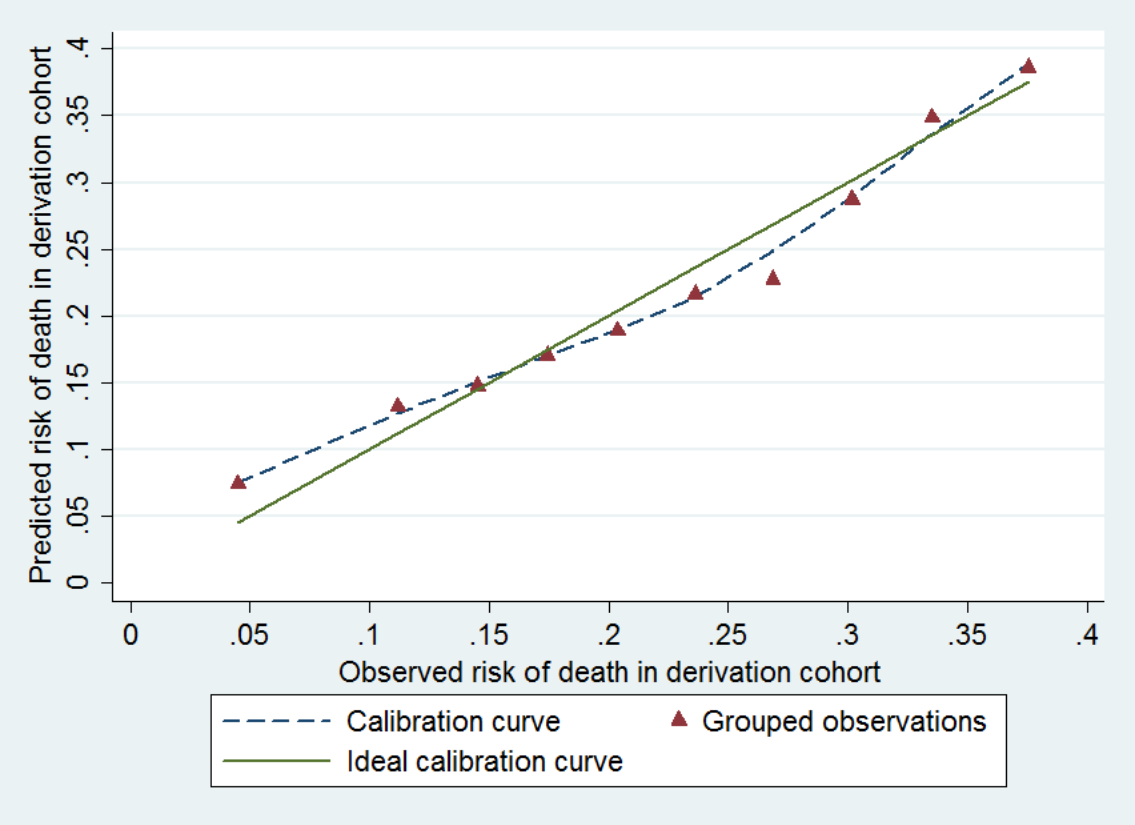
Supplemental Figure 3. Calibration curve in the PLR model for stroke in the derivation cohort



Supplemental Figure 4. Calibration curve in the PLR model for major bleeding in the derivation cohort



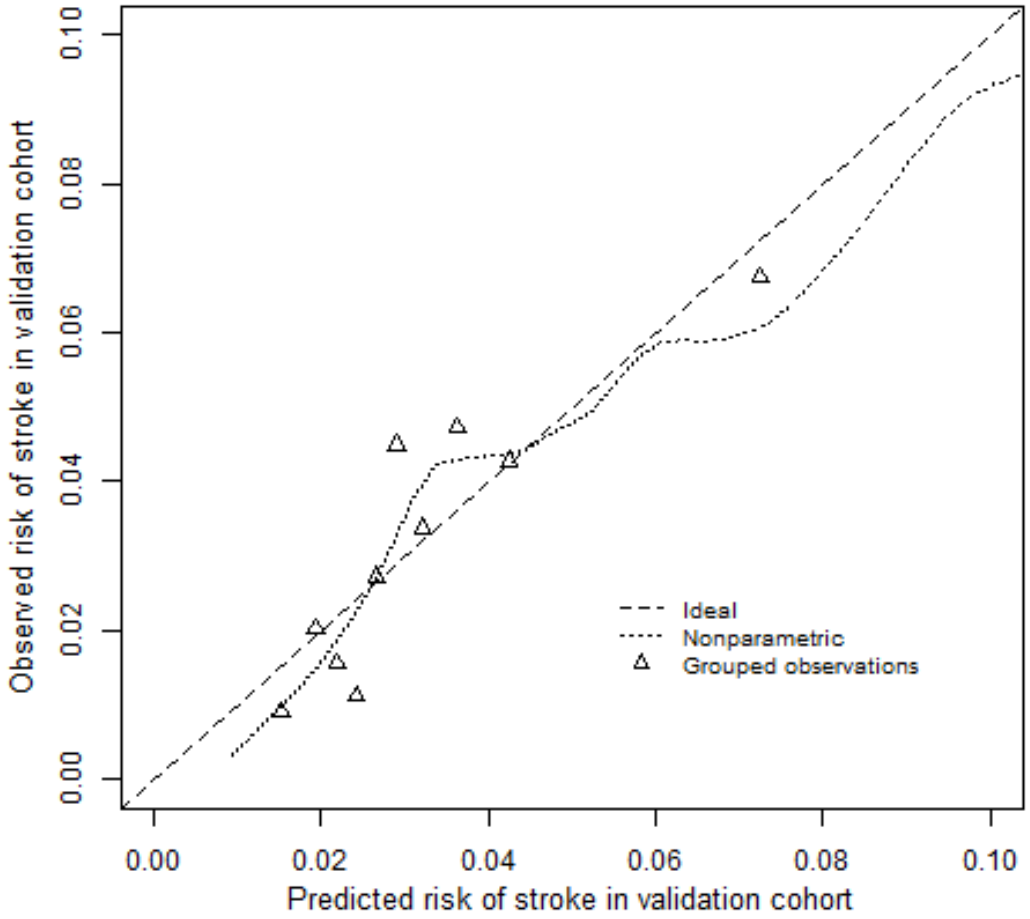
Supplemental Figure 5. Calibration curve in the Cox model for death in the derivation cohort



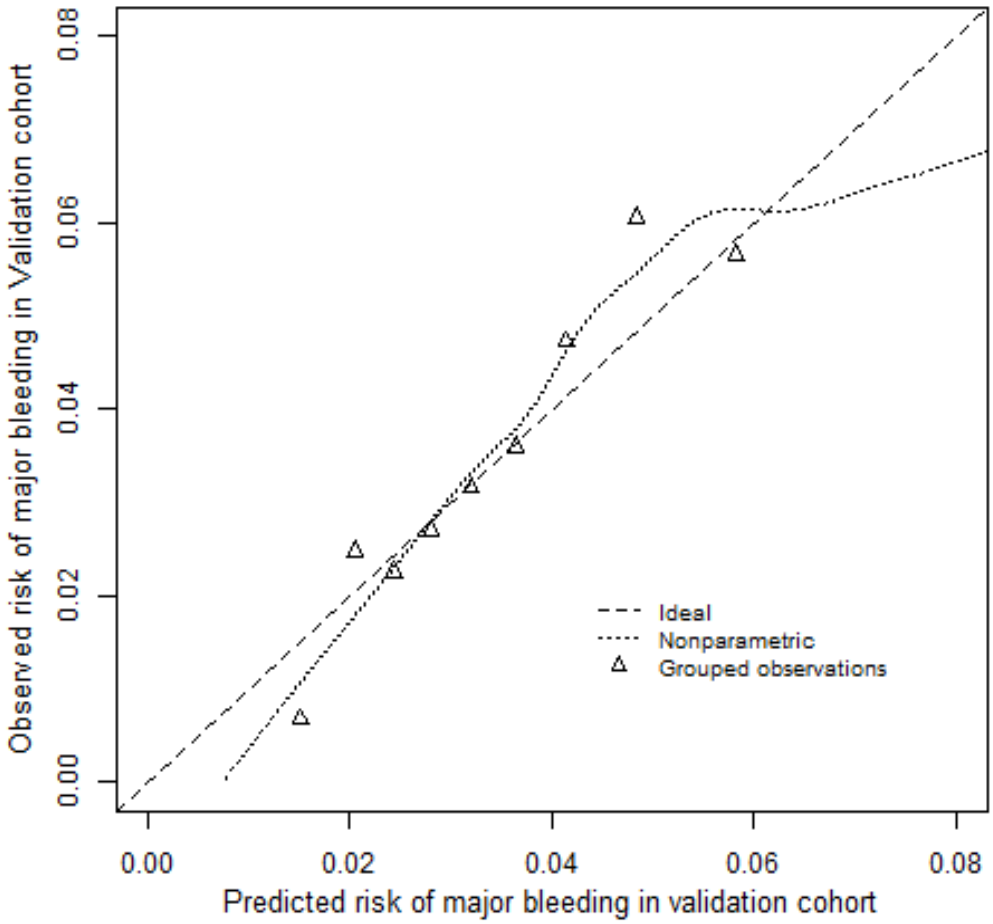
Supplemental Table 4. Updates of the models' intercepts and the regression coefficients for external validation in the KPCO-Ⅱ cohort

Recalibration of model	Update of the PLR model		Update of the Cox model
	Stroke (n=142) vs. neither event	Major bleeding (n=173) vs. neither event	All-cause death (n=992) vs. survivors
Calibration intercept (95% CI, p-value)	-2.74 (-2.98 to -2.50), < 0.001	-3.58 (-3.79 to -3.37), < 0.001	Not specified
Calibration slope (95% CI, p-value)	0.70 (0.49 to 0.91), < 0.001	0.66 (0.47 to 0.85), < 0.001	1.02 (0.95-1.09), < 0.001

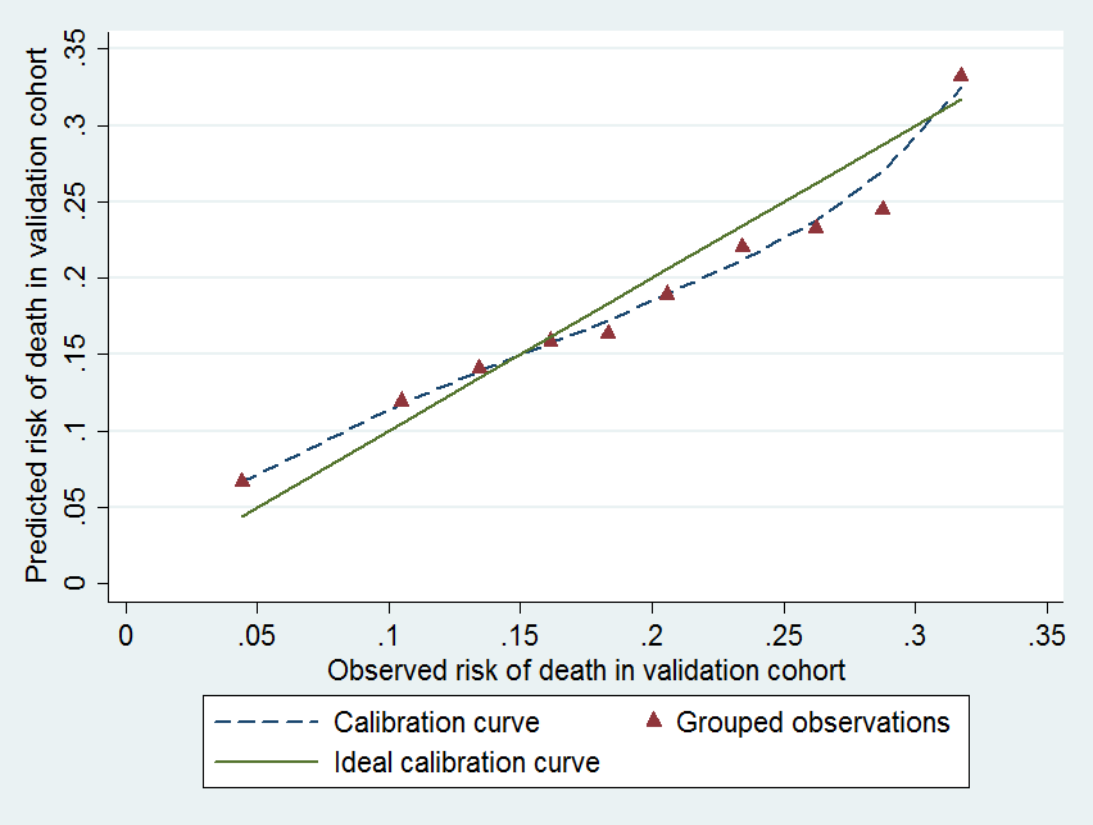
Supplemental Figure 6. Calibration curve in the PLR model for stroke in the validation cohort



Supplemental Figure 7. Calibration curve in the PLR model for major bleeding in the validation cohort



Supplemental Figure 8. Calibration curve in the Cox model for death in the validation cohort



Supplemental Table 5. Hypertensive drugs as surrogates for hypertension diagnosis

Group	Drug list
Angiotensin-converting enzyme inhibitors	captopril, benazepril, enalapril, lisinopril, fosinopril, ramipril, guinapril, tranolopril, perinodopril
Angiotensin II receptor blockers	candesartan, eprosartan, irbesartan, losartan, olmesartan, telmisartan, valsartan
Thiazides	chlorthiazide, benzthiazide, trichlormethiazide, hydrochlorthiazide, hydroflumethiazide, chlorthalidone, indapamide, methylclothiazide, polythiazide, metolazone
Beta-blockers	alprenolol, carteolol, levobunolol, mepindolol, nadolol, oxprenolol, penbutolol, pindolol, propranolol, sotalol, timolol, acebutalol, atenolol, betaxolol, bisoprolol, esmolol, metoprolol, nebivolol, carvedilol, celiprolol, labetolol
Calcium channel blockers	amlodipine, felodipine, nicardipine, nifedipine, nimodipine, nisoldipine, nitrendipine, lacidipine, verapamil, diltiazem
Others	clonidine, methyldopa, hydralazine

Supplemental Table 6. Results for effect of hypertension in the PLR and Cox model using different data on hypertension in the KPCO-I cohort*

Effect of hypertension	**PLR model** (expressed as OR with 95% CI, p-value)		**Cox model** (expressed as HR with 95% CI, p-value)
	Stroke vs. neither event	**Major bleeding vs. neither event**	**Death vs. survival**
Either ICD-9-CM codes or hypertensive drugs	0.88 (0.56-1.39), 0.587	0.94 (0.66-1.33), 0.712	0.76 (0.66-0.85), <0.001
ICD-9-CM Codes only	0.80 (0.55-1.17), 0.256	0.84 (0.64-1.09), 0.195	0.75 (0.67-0.85), <0.001
Both ICD-9-CM codes and hypertensive drugs	0.86 (0.66-1.12), 0.269	0.97 (0.67-1.42), 0.893	0.76 (0.67-0.86), <0.001

*The three approaches produced the same predictors included in the models and extremely similar coefficients; coefficients for the other predictors not shown here.

Supplemental Table 7. Sensitivity analysis leaving hypertension out of the PLR model for stroke and major bleeding in the KPCO-I cohort

Predictors	Stroke vs. neither event (OR with 95%CI, p-values)[1]	Major bleeding vs. neither event (OR with 95%CI, p-values)[2]
Intercept: coefficient β , p-value	-3.89, <0.001	-4.02, <0.001
Age[3]	1.02 (1.00-1.03), 0.015	1.02 (1.01-1.03), 0.001
Female	1.50 (1.05-2.15), 0.027	0.73 (0.56-0.94), 0.014
Warfarin	0.98 (0.69-1.40), 0.931	1.87 (1.45-2.42), < 0.001
Other cerebrovascular disease	4.79 (2.44-9.42), <0.001	1.38 (0.60-3.19), 0.454
Congestive heart failure	0.79 (0.44-1.40), 0.412	1.57 (1.12-2.19), 0.008
Diabetes	1.11 (0.71-1.75), 0.650	1.19 (0.88-1.62), 0.261
Prior major bleeding	1.13 (0.50-2.54), 0.767	1.49 (0.87-2.55), 0.144
Prior stroke	2.02 (1.15-3.54), 0.014	0.71 (0.39-1.30), 0.262
Renal disease	1.34 (0.74-2.43), 0.339	1.47 (0.98-2.21), 0.065
Concurrent use of antibiotics	0.98 (0.63-1.50), 0.912	1.81 (1.38-2.36), < 0.001
Concurrent use of antiplatelets	1.70 (1.05-2.74), 0.032	1.58 (1.09-2.27), 0.016
Concurrent use of gastrointestinal medications	0.83 (0.53-1.31), 0.425	1.76 (1.34-2.31), < 0.001

[1] AUC for stroke versus neither event: 0.69 (0.65-0.74); Hosmer-Lemeshow test statistics (p-value): 8.89 (0.352)

[2] AUC for major bleeding versus neither event: 0.71 (0.68-0.75); Hosmer-Lemeshow test statistics (p-value): 10.02 (0.264)

[3] Used as per one-year change

Supplemental Table 8. Sensitivity analysis leaving hypertension out of the Cox model for death in the KPCO-I cohort

Predictors	All-cause death (n=1194)[1]	
	HR (95% CI)	P-value
Age[2]	1.06 (1.06-1.07)	<0.001
Warfarin	0.52 (0.46-0.58)	< 0.001
Anemia	1.85 (1.56-2.19)	< 0.001
Other cerebrovascular disease[3]	1.68 (1.21-2.34)	0.002
Congestive heart failure	1.46 (1.26-1.71)	< 0.001
Diabetes	1.41 (1.23-1.62)	< 0.001
Prior major bleeding	1.37 (1.07-1.75)	0.014
Malignancy[2]	1.82 (1.56-2.14)	<0.001
Concurrent use of antifungals	1.57 (1.12-2.19)	0.008
Concurrent use of antidepressants	1.21 (1.03-1.42)	0.023

[1] C-index = 0.74 (0.73-0.76); Groennesby and Borgan test statistic (p-value): 14.05 (0.120)

[2] Used as per one-year change; [3] Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin

Chapter 4

Risk factors for and prediction of mortality in critically ill medical–surgical patients receiving heparin thromboprophylaxis

Guowei Li[1], Lehana Thabane[1,2], Deborah J. Cook[1,2,3], Renato D. Lopes[4], John C. Marshall[5,6], Gordon Guyatt[1,3], Anne Holbrook[1,2,3], Noori Akhtar-Danesh[1,7], Robert A. Fowler[5,8], Neill K. J. Adhikari[5,8], Rob Taylor[9], Yaseen M. Arabi[10], Dean Chittock[11], Peter Dodek[12], Andreas P. Freitag[3], Stephen D. Walter[1], Diane Heels-Ansdell[1] and Mitchell A. H. Levine[1,2,3*]

Abstract

Background: Previous studies have suggested that prediction models for mortality should be adjusted for additional risk factors beyond the Acute Physiology and Chronic Health Evaluation (APACHE) score. Our objective was to identify risk factors independent of APACHE II score and construct a prediction model to improve the predictive accuracy for hospital and intensive care unit (ICU) mortality.

Methods: We used data from a multicenter randomized controlled trial (PROTECT, Prophylaxis for Thromboembolism in Critical Care Trial) to build a new prediction model for hospital and ICU mortality. Our primary outcome was all-cause 60-day hospital mortality, and the secondary outcome was all-cause 60-day ICU mortality.

Results: We included 3746 critically ill non-trauma medical–surgical patients receiving heparin thromboprophylaxis (43.3 % females) in this study. The new model predicting 60-day hospital mortality incorporated APACHE II score (main effect: hazard ratio (HR) = 0.97 for per-point increase), body mass index (BMI) (main effect: HR = 0.92 for per-point increase), medical admission versus surgical (HR = 1.67), use of inotropes or vasopressors (HR = 1.34), acetylsalicylic acid or clopidogrel (HR = 1.27) and the interaction term between APACHE II score and BMI (HR = 1.002 for per-point increase). This model had a good fit to the data and was well calibrated and internally validated. However, the discriminative ability of the prediction model was unsatisfactory (C index < 0.65). Sensitivity analyses supported the robustness of these findings. Similar results were observed in the new prediction model for 60-day ICU mortality which included APACHE II score, BMI, medical admission and invasive mechanical ventilation.

Conclusion: Compared with the APACHE II score alone, the new prediction model increases data collection, is more complex but does not substantially improve discriminative ability.

Trial registration: ClinicalTrials.gov Identifier: NCT00182143

Keywords: Prediction model, Critical care, APACHE, Intensive care unit, Mortality

Background

Mortality rates in critically ill patients are substantial, ranging from 5 to 40 %, depending on case mix [1–3]. Predicting mortality in critically ill patients is challenging, but can be helpful for general counseling, triaging, treatment decisions and end of life discussions [4, 5].

The Acute Physiology and Chronic Health Evaluation (APACHE) prognostic scoring system is a well-established, validated tool for assessing the severity of illness and predicting hospital mortality using data obtained in the first 24 h of ICU admission [6–11]. To increase

*Correspondence: levinem@mcmaster.ca
[2] Centre for Evaluation of Medicines, St. Joseph's Healthcare Hamilton, McMaster University, 25 Main St. West, Suite 2000, 20th Floor, Hamilton, ON L8P 1H1, Canada
Full list of author information is available at the end of the article

predictive accuracy, APACHE has developed four generations of models [12–15]. Nevertheless, there may be wide variation and limited validation in the ability of APACHE system to predict mortality in different countries and populations [16–20]. Siontis et al. reported a median AUC (the area under the receiver operating characteristic curve) of 0.77 for APACHE II model after conducting a systematic evaluation of predictive tools for all-cause mortality in critically ill patients [21]. Furthermore, the updated APACHE III and IV models include substantially more variables than APACHE II, with a correspondingly increased data collection burden [7, 22].

Advances in prognostic science have identified additional risk factors for mortality for critically ill patients that are independent of measures of illness severity, such as body mass index (BMI) [23, 24] and sex [25–27]. Moreover, some evidence suggests that prediction models for mortality should be adjusted for the use of vasopressors [28], prothrombin index [29, 30] and platelet count [31, 32]. Therefore, given the imperfect accuracy of the APACHE system and other potential risk factors for death, we aimed to identify risk factors independent of APACHE II score and construct and validate a new mortality prediction model that would combine the APACHE II score with these additional factors. The primary objective of this study was to improve the accuracy of a prediction model for 60-day hospital mortality in critically ill medical–surgical patients, based on the data from a multicenter randomized controlled trial, PROTECT (Prophylaxis for Thromboembolism in Critical Care Trial). Our secondary objective was to construct a prediction model for 60-day ICU mortality.

Methods

In this study, we followed the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement [33] to report the prediction model including model development, model performance and model validation.

Patients and settings

PROTECT (ClinicalTrials.gov Identifier: NCT00182143) was an international randomized controlled trial (RCT) that was conducted in 67 ICUs in academic and community hospitals from 2006 to 2010 in Canada, Australia, Brazil, Saudi Arabia, the USA and the UK, as described elsewhere [34]. The trial compared the effect of unfractionated heparin (UFH) 5000 IU twice daily versus the low molecular weight heparin (LMWH) dalteparin 5000 IU once daily plus once-daily placebo on the primary outcome of proximal leg deep vein thrombosis.

Non-trauma medical–surgical critically ill patients were enrolled if they were at least 18 years of age, weighed ≥45 kg and were expected to remain in the ICU for at least 3 days. Exclusion criteria were: admission diagnoses of major trauma, neurosurgery or orthopedic surgery, uncontrolled hypertension (systolic blood pressure >180 mm Hg or diastolic blood pressure >110 mm Hg) for at least 12 h, major bleeding within the last week unless definitively treated, hemorrhagic stroke, coagulopathy (international normalized ratio >2 times upper limit of normal or activated partial thromboplastin time >2 times the upper limit of normal), severe thrombocytopenia (platelet count $<75 \times 10^9$/L), need for therapeutic anticoagulation, heparin administration in the ICU for at least 3 days, contraindication to heparin or blood products, pregnancy, life-support limitation, life expectancy ≤7 days or enrollment in another related trial [34, 35]. All patients, families, clinicians, research personnel and the trial biostatistician were blind to treatment allocation. Patients were followed up to death or hospital discharge.

Outcome measures

During the trial follow-up, the vital status was documented in the ICU and in hospital. In this study, the primary outcome was 60-day hospital mortality. Patients survived longer than 60 days in hospital or discharged from hospital were censored. The secondary outcome was 60-day ICU mortality, and patients survived long than 60 days in ICU or discharged from the ICU were censored.

Potential predictors

Based on the data recorded in PROTECT and our a priori plan, potential risk factors for death included baseline variables (APACHE II score, sex, BMI, history of malignancy, type of admission and diagnosis of sepsis on admission), use of UFH and interventions within the first 24 h of ICU admission (use of inotropes or vasopressors, invasive mechanical ventilation, dialysis and pharmacologic cointerventions). The APACHE II score [13] has three parts: an acute physiology score (up to 60 points), an age point (0–6) and a chronic health score (0–5). The acute physiology score is composed of 12 physiologic variables: creatinine (0–8 points); Glasgow Coma Scale (0–12 points); ten other variables including temperature, mean arterial pressure, heart rate, respiratory rate, oxygenation, arterial pH, serum sodium, potassium, hematocrit and white blood cell count (0–4 points each). The maximum total APACHE II score is 71 points, and a higher score indicates a higher predicted probability of death. The type of admission was categorized as either surgical or medical. Pharmacologic cointerventions included the use of a statin, and acetylsalicylic acid or clopidogrel.

Statistical analyses

In this study, all analyses were conducted using STATA version 12 (Stata Corp., College Station, TX, USA). Data were summarized using the mean and standard deviation (SD), or median and interquartile range (IQR) or frequency and percentages. Comparisons between the patients who died and survived for the duration of hospital stay were made by using Student's *t* test for continuous variables and Chi-square test for categorical variables, respectively. If <10 % of observations on a variable were missing, we imputed the missing values using the mean or median. If ≥10 % of data were missing, multiple imputations were performed, assuming they were missing at random [36].

Identification of risk factors independent of APACHE II score and model development

To identify risk factors independent of APACHE II score, data were first randomly split into a training (derivation) set and a validation set stratifying by participating trial centers. The derivation set and validation set had an approximately equal sample size. In the derivation set, to avoid multicollinearity, we pruned the candidate predictors of those with a variance inflation factor (VIF) of no less than 4 [37, 38]. Cox proportional hazards regression was conducted to examine associations with death using the backward elimination approach [37], after adjustment for the APACHE II score, with a two-sided alpha value of 0.05. Hazard ratios (HRs) were used to quantify the relationship between risk factors and death. Both a statistical test of proportional hazards assumption and a graphical examination using Schoenfeld residuals were performed to test the proportional hazards assumption of the Cox regression models [39].

In the derivation set, the new prediction model for 60-day hospital mortality was constructed by combining the APACHE II score and the other risk factors identified above into a Cox regression model. Additionally, all the two-way interactions between the predictors in the new prediction model were tested. Significant interactions with an a priori alpha value of 0.05 were then added into the model to finalize the prediction model.

For 60-day ICU mortality, identification of risk factors independent of APACHE II score and construction of a new prediction model were performed in the whole dataset following the same process.

Model performance

For succinctness, we defined three models for hospital and ICU mortality in this study: Model 1 which included the APACHE II score only; Model 2 that included the other risk factors only; and Model 3, as the new prediction model, which combined APACHE II score and the other risk factors. To assess the calibration of all the three models for 60-day hospital mortality in the derivation set, we calculated the standardized mortality ratio (SMR) by dividing the observed death risk by the predicted mortality. To obtain the 95 % confidence intervals (CIs) for SMRs, first we treated the observed mortality as a Poisson variable, and then divided its 95 % confidence limits by the predicted mortality [40]. For Model 1 and Model 3, we also compared and plotted the predicted and observed risks of death across each 10th of observed risk [41], in which the observed risk was obtained from the Kaplan–Meier product-limit estimate.

Model goodness-of-fit was evaluated using a Gronnesby and Borgan test with ten groups based on the predicted risk score, where a nonsignificant result indicated no evidence of lack of fit to the data [42]. The Akaike information criterion (AIC) was used to evaluate and compare the goodness-of-fit between the three models; a smaller AIC value indicated a better model [43]. The likelihood ratio test was also performed for model comparison. To measure discrimination, we calculated a Harrell's *C* index for each model [37, 44].

For 60-day ICU mortality, performance of the three models was assessed and compared using the whole dataset.

Model validation for hospital mortality

We used the validation set to assess the internal validation of all the three models [45]. The evaluation of calibration, goodness-of-fit and discrimination was again performed in the validation set [45, 46].

Sensitivity analyses

To assess the robustness of findings, we performed a sensitivity analysis by using restricted cubic splines for continuous predictors in the new model [37]. Another sensitivity analysis was conducted using data for 30-day hospital and 30-day ICU mortality, and 90-day hospital and 90-day ICU mortality.

Exploratory analysis for hospital mortality

We applied and compared Model 3 and Model 1 in different countries for 60-day hospital mortality in the whole dataset, as an exploratory analysis. Model performance was assessed separately in Canada, Saudi Arabia and Brazil, USA and UK, and Australia.

Results

Baseline characteristics of participants

There were 3746 patients included for analyses. The mean age at baseline was 61.4 (SD: 16.5) years, and 43.3 % were females. The median survival of the 588 (15.7 %) patients who died in the ICU was 10 days. The median survival

of the 873 (23.3 %) patients who died during the hospital stay was 14 days.

The data were randomly split into a derivation set (n = 1891) and a validation set (n = 1855). Figure 1 shows Kaplan–Meier survival curves of 60-day hospital mortality in the derivation and validation sets, with no evidence of significant difference between the two sets (p value = 0.94 for log-rank test). Table 1 compares the baseline characteristics between the survivors and non-survivors in the derivation and validation sets. In the derivation set, 22.6 % of participants (n = 428) died during the whole follow-up period and their median survival time was 14 days (IQR 7.5–28). The median follow-up for survivors (n = 1463) was 18 days (IQR 11–33). Non-survivors were significantly older than survivors (67.7 vs. 59.3 years). The survivors had significantly lower APACHE II scores but higher BMI than non-survivors (p value <0.001). There were more patients receiving UFH in non-survivors (54.7 %) than in survivors (49.1 %). More non-surviving patients were admitted to ICU with the diagnoses of sepsis and medical reasons (p value <0.001). Non-survivors were significantly more likely to receive inotropes or vasopressors, invasive mechanical ventilation, dialysis and acetylsalicylic acid or clopidogrel within the first 24 h of ICU admission (p value <0.05). Similar comparisons were also found in the validation set between non-survivors and survivors, except for the proportions of patients receiving UFH, invasive mechanical ventilation and dialysis, and the percentages of patients with malignancy and medical admission (Table 1).

Model construction

Table 2 shows the predictors and their HRs included in the new model (Model 3) for 60-day hospital and 60-day ICU mortality. Based on the derivation set, BMI, medical admission, use of inotropes or vasopressors and acetylsalicylic acid or clopidogrel were significant risk factors for 60-day hospital mortality independently of APACHE II score; all of them except BMI increased the risk of hospital death. Model 3 for hospital mortality included APACHE II score (main effect: HR = 0.97, 95 % CI 0.92–1.02 for per-point increase), BMI (main effect: HR = 0.92, 95 % CI 0.88–0.97 for per-point increase), medical admission (HR = 1.67, 95 % CI 1.29–2.17), use of inotropes or vasopressors (HR = 1.34, 95 % CI 1.10–1.65), acetylsalicylic acid or clopidogrel (HR = 1.27, 95 % CI 1.02–1.59) and the interaction term between APACHE II score and BMI (HR = 1.002, 95 % CI 1.000–1.004 for per-point increase) (Table 2).

Significant risk factors for 60-day ICU mortality independent of APACHE II score were BMI, medical admission and invasive mechanical ventilation. Model 3 for ICU mortality included APACHE II score, BMI, medical admission and invasive mechanical ventilation, with a HR of 1.04, 0.98, 1.39 and 0.75, respectively. No significant interaction terms were identified for Model 3 (Table 2).

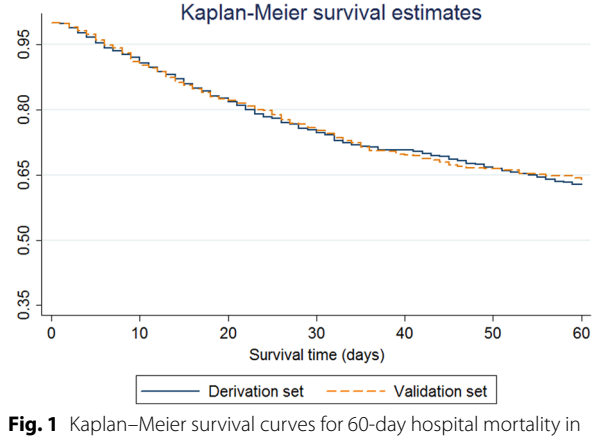


Fig. 1 Kaplan–Meier survival curves for 60-day hospital mortality in derivation and validation sets

Model performance

Results and comparison of the three models for 60-day hospital mortality are shown in Table 3. In the derivation set, the goodness-of-fit test indicated no evidence of lack of fit to the data for Model 1 (p value = 0.68) for hospital mortality. However, the discriminative ability of Model 1 was poor (C index = 0.58). No evidence for the inaccurate overall prediction of mortality by Model 1 was found, given that the SMR was not significantly different from 1 (SMR = 1.003, 95 % CI 0.959–1.050) (Table 3). Figure 2a displays predicted and observed hospital mortality in the derivation set across each 10th of the observed risk of death for Model 1, indicating Model 1 was well calibrated. Similarly, Model 2 was a good fit and well calibrated in the derivation set, but its discriminative power was not high (C index = 0.62). Model 3 had a C index of 0.64 and a SMR of 1.006 (95 % CI 0.961–1.052) (Table 3). The difference in C indices between Model 3 and Model 1 was significant (p value <0.001). Figure 2b shows predicted versus observed hospital mortality in the derivation set, which justified the calibration of Model 3. The smallest AIC was observed in Model 3, indicating that Model 3 performed better than the other two models (Table 3). Likelihood ratio test also implied that Model 3 was a better fit than Model 1 and Model 2 (p values <0.001).

When the models were applied to the validation set, findings were unchanged. All the three models were well calibrated (Table 3; Fig. 3a for Model 1 and Fig. 3b for Model 3, respectively); nevertheless, their discriminative

Table 1 Baseline characteristics of survivors and non-survivors in hospital in derivation and validation datasets

Characteristics	Derivation set ($n = 1891$)			Validation set ($n = 1855$)		
	Survivors[a] ($n = 1463$)	Non-survivors[b] ($n = 428$)	*p* value	Survivors[c] ($n = 1410$)	Non-survivors[d] ($n = 445$)	*p* value
Age (year): mean (SD)	59.3 (16.92)	67.7 (14.58)	<0.001[e]	59.6 (16.07)	68.3 (14.40)	<0.001[e]
Gender: *n* (%)						
Male	830 (56.93)	231 (54.10)	0.301[f]	786 (56.22)	266 (59.91)	0.172[f]
Female	628 (43.07)	196 (45.90)		612 (43.78)	178 (40.09)	
BMI (kg/m^2): mean (SD)	28.5 (7.69)	27.1 (6.93)	<0.001[e]	28.9 (8.39)	27.1 (6.78)	<0.001[e]
Use of thromboprophylaxis: *n* (%)						
Unfractionated heparin	718 (49.08)	234 (54.67)	0.042[f]	696 (49.36)	225 (50.56)	0.659[f]
Dalteparin	745 (50.92)	194 (45.33)		714 (50.64)	220 (49.44)	
APACHE II score: mean (SD)	20.8 (7.61)	24.1 (7.59)	<0.001[e]	20.6 (7.59)	24.5 (7.84)	<0.001[e]
History of malignancy: *n* (%)	50 (3.43)	21 (4.92)	0.155[f]	47 (3.36)	32 (7.21)	<0.001[f]
Medical admission: *n* (%)	1086 (74.23)	353 (82.48)	<0.001[f]	1046 (74.18)	346 (77.75)	0.129[f]
Diagnosis of sepsis on admission: *n* (%)	208 (14.27)	90 (21.08)	<0.001[f]	177 (12.66)	74 (16.67)	0.032[f]
Intervention within the first 24 h on admission: *n* (%)						
Inotropes or vasopressors	614 (42.20)	241 (56.44)	<0.001[f]	579 (41.42)	243 (54.73)	<0.001[f]
Invasive mechanical ventilation	1207 (82.96)	373 (87.35)	0.030[f]	1149 (82.19)	380 (85.59)	0.097[f]
Dialysis	57 (3.92)	45 (10.54)	<0.001[f]	87 (6.22)	39 (8.78)	0.063[f]
Acetylsalicylic acid or clopidogrel	290 (19.93)	113 (26.46)	0.004[f]	289 (20.49)	112 (25.23)	0.029[f]
Statin	169 (11.62)	62 (14.52)	0.108[f]	199 (14.23)	59 (13.29)	0.617[f]

[a] Median follow-up: 18 days; interquartile range (IQR): 11–33 days

[b] Median survival time: 14 days; IQR 7.5–28 days

[c] Median follow-up: 19 days; IQR 11–37 days

[d] Median survival time: 15 days; IQR 7–29 days

[e] Student's *t* test

[f] Chi-square test

Table 2 Predictors for 60-day hospital mortality in the derivation dataset and for 60-day ICU mortality in the whole dataset

Predictors	Hospital mortality ($n = 1891$)[a]		ICU mortality ($n = 3746$)[b]	
	HR (95 % CI)	*p* value	HR (95 % CI)	*p* value
BMI	0.92 (0.88–0.97)	0.003	0.98 (0.96–0.99)	<0.001
Medical admission	1.67 (1.29–2.17)	<0.001	1.39 (1.11–1.72)	0.003
Inotropes or vasopressors	1.34 (1.10–1.65)	0.005	–[c]	–[c]
Acetylsalicylic acid or clopidogrel	1.27 (1.02–1.59)	0.035	–[c]	–[c]
APACHE II score	0.97 (0.92–1.02)	0.241	1.04 (1.03–1.05)	<0.001
APACHE II score*BMI	1.002 (1.000–1.004)	0.038	–[c]	–[c]
Invasive mechanical ventilation	–[d]	–[d]	0.75 (0.58–0.97)	0.027

[a] There were 390 60-day deaths in hospital in derivation cohort

[b] There were 573 60-day deaths in ICU in the whole cohort

[c] Not included in the model for ICU mortality; no interaction term in the model for ICU mortality

[d] Not included in the model for hospital mortality

power was not satisfactory (Table 3). Results from AIC and likelihood ratio tests presented that Model 3 was better than Model 1 and Model 2.

Table 3 also displays results for 60-day ICU mortality using the whole dataset. The SMR was 1.003 and 1.004 for Model 1 and Model 3, respectively. Figure 4a and b also

Table 3 Comparing three models in model performance for 60-day hospital mortality in the derivation and validation dataset and for 60-day ICU mortality in the whole dataset

Model performance	Goodness-of-fit		*C* index	SMR (95 % CI)
	p value[a]	AIC		
Hospital mortality—derivation set ($n = 1891$)				
Model 1 (including APACHE II scores only)	0.68	5704	0.58	1.003 (0.959–1.050)
Model 2 (including the other risk factors only)[b]	0.16	5359	0.62	1.002 (0.956–1.049)
Model 3 (including both the other risk factors and APACHE II scores)[c]	0.90	5329	0.64	1.006 (0.961–1.052)
Hospital mortality—validation set ($n = 1855$)				
Model 1 (including APACHE II scores only)	0.37	5912	0.60	0.933 (0.890–0.978)
Model 2 (including the other risk factors only)[b]	0.80	5765	0.59	1.019 (0.972–1.067)
Model 3 (including both the other risk factors and APACHE II scores)[c]	0.88	5567	0.64	1.011 (0.966–1.060)
ICU mortality—the whole set ($n = 3746$)				
Model 1 (including APACHE II scores only)	0.75	8180	0.61	1.003 (0.972–1.036)
Model 2 (including the other risk factors only)[d]	0.24	7821	0.58	1.001 (0.969–1.034)
Model 3 (including both the other risk factors and APACHE II scores)[e]	0.74	7778	0.64	1.004 (0.972–1.038)

AIC Akaike information criterion, *SMR* standardized mortality ratio

[a] Based on Groennesby and Borgan test

[b] The other risk factors included BMI, medical admission, inotropes or vasopressors and acetylsalicylic acid or clopidogrel

[c] Model 3 consisted of BMI, medical admission, inotropes or vasopressors, acetylsalicylic acid or clopidogrel, APACHE II score and the interaction between BMI and APACHE II score

[d] The other risk factors included BMI, medical admission and invasive mechanical ventilation

[e] Model 3 consisted of BMI, medical admission, invasive mechanical ventilation and APACHE II score

support the calibration of Model 1 and Model 3, respectively. The *C* index was not high, with a discriminative value of 0.61 and 0.64 for Model 1 and Model 3, respectively. No significant difference in *C* indices between Model 3 and Model 1 was observed (p value $= 0.16$).

Results from sensitivity analysis using restricted cubic splines for BMI, APACHE II score and the interaction between them displayed similar findings from Model 3 for 60-day hospital mortality, where the interior knots were located on 25 and 30 for BMI, and the medians for APACHE II score (21) and the interaction term (569), respectively (Table 4). Findings were also in good agreement with Model 3 for 60-day ICU mortality when restricted cubic splines were used for BMI and APACHE II score (Table 4). Similar results of model construction and model performance were observed in another sensitivity analysis limiting data to 30-day hospital and 30-day ICU mortality (Appendix Tables 6, 7) and restricting data to 90-day hospital and 90-day ICU mortality (Appendix Tables 8, 9).

Exploratory analysis

An exploratory analysis was conducted by country for hospital mortality using the whole dataset (Table 5). Similar model performance was observed in different countries using Model 1 and Model 3. However, evidence indicated that Model 1 may under-predict risk of 60-day hospital death for patients in Saudi Arabia and Brazil (SMR = 1.155, 95 % CI 1.054–1.263).

Discussion

Main findings

Based on the data from an international thromboprophylaxis trial, we identified risk factors other than APACHE II score which predicted 60-day hospital mortality and 60-day ICU mortality. We constructed a new prediction model for mortality in critically ill patients receiving thromboprophylaxis. The new model was a good fit, well calibrated and internally validated. Results from sensitivity analyses supported the robustness of findings. However, the discriminative ability of the prediction model was not satisfactory.

In this study, we identified that higher BMI was significantly related to decreased risk of hospital and ICU mortality (Table 2), which was congruent with previous studies [47–49]. The potentially protective effect of increased BMI on survival has been termed the obesity paradox or reverse epidemiology [50], but whether the observed association is causative remains unresolved [49, 51, 52]. It has been postulated that higher body weight affords nutritional reserves that increase the chance of survival when patients are critically ill [53].

Medical admission was also found to be a significant independent risk factor for hospital and ICU death. Patients admitted for medical reasons may have more serious chronic morbidities not fully accounted for by APACHE chronic conditions, or have poorer prognoses when admitted to the ICU compared to those patients

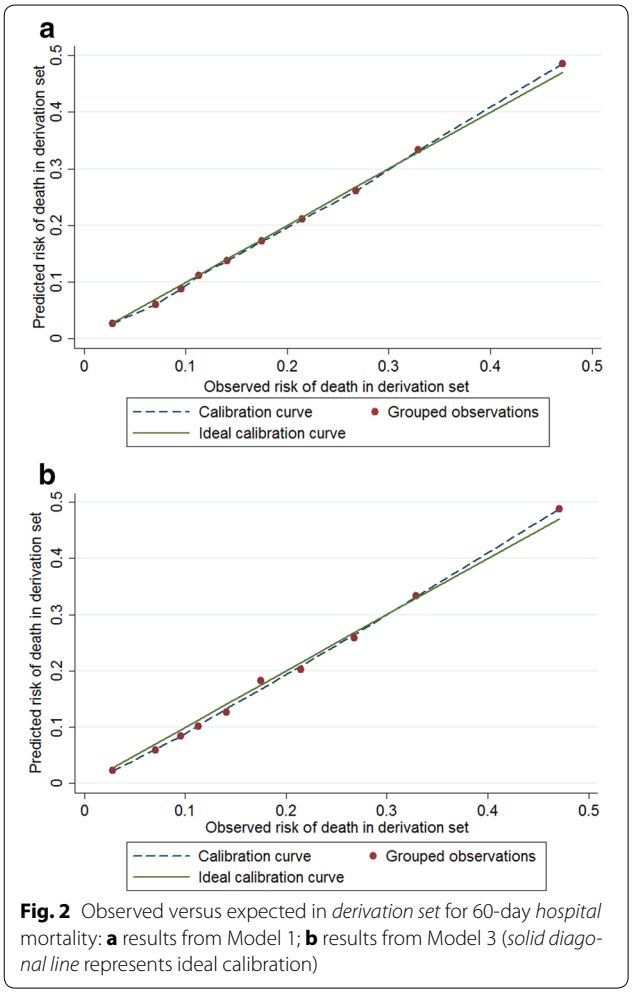


Fig. 2 Observed versus expected in *derivation set* for 60-day *hospital* mortality: **a** results from Model 1; **b** results from Model 3 (*solid diagonal line* represents ideal calibration)

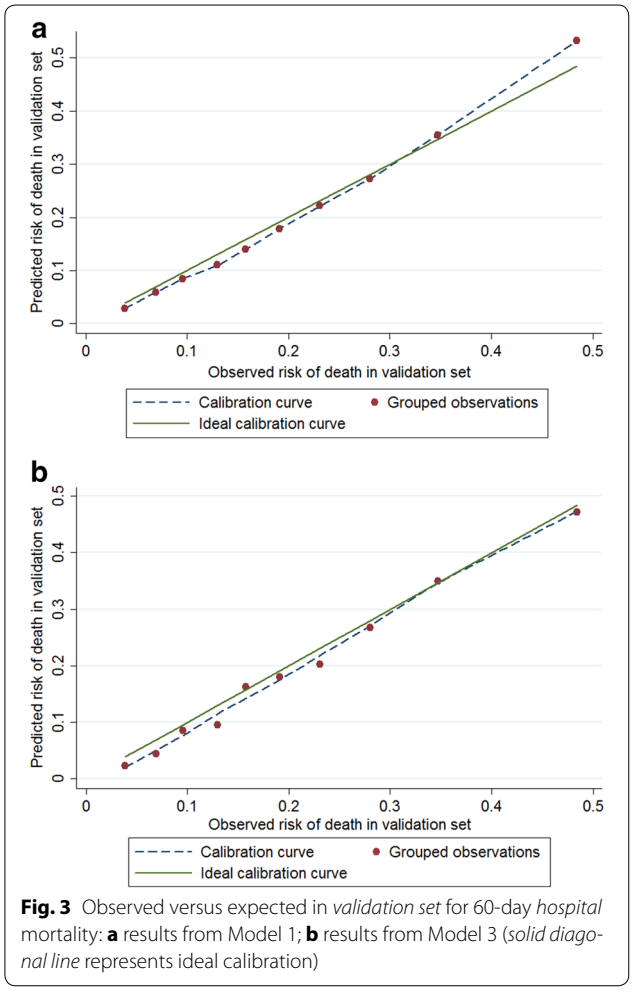


Fig. 3 Observed versus expected in *validation set* for 60-day *hospital* mortality: **a** results from Model 1; **b** results from Model 3 (*solid diagonal line* represents ideal calibration)

selected for surgery. Also, we found that some interventions within the first 24 h of ICU admission such as use of inotropes or vasopressors, and acetylsalicylic acid or clopidogrel, were associated with increased risk of hospital mortality, reflecting more severe illness. However, invasive mechanical ventilation within the first 24 h on admission was associated with 25 % decreased risk of ICU death (Table 2). Evidence suggests that not using invasive mechanical ventilation could have negative effects on outcome by postponing necessary intubation; therefore, early initiation of invasive mechanical ventilation may be related to decreased risk of death [54, 55].

Implications of the study

Given the previously acknowledged limited predictive accuracy of the APACHE II system for mortality, we sought to build a new prediction model for mortality for critically ill medical–surgical patients. In this study, the model including APACHE II score only (Model 1) had surprisingly low discriminative ability (Tables 3, 5), which has been documented previously [21]. The model that combined additional baseline characteristics and early ICU interventions may better assess patients' illness severity and thus improve the estimated risk of mortality, compared with the APACHE II score alone. Nevertheless, though the prediction model had a significantly higher *C* index than APACHE II score in predicting risk of hospital mortality, adding more information such as BMI, medical admission and early pharmacologic interventions increased the discriminative accuracy to only a small extent (Table 3). The simplicity of the APACHE II score is a major reason why it remains the most commonly used severity scoring system globally in clinical practice as well as health research [4]. Compared with the APACHE II score alone, the utilization of a new model which increases data collection, is more complex but does not substantially improve discriminative ability. Therefore, the use of the new model would be limited to situations where a clinician or health services investigator was sufficiently dissatisfied with the APACHE II and was requiring an even minimally better model to predict risk of death in critically ill patients.

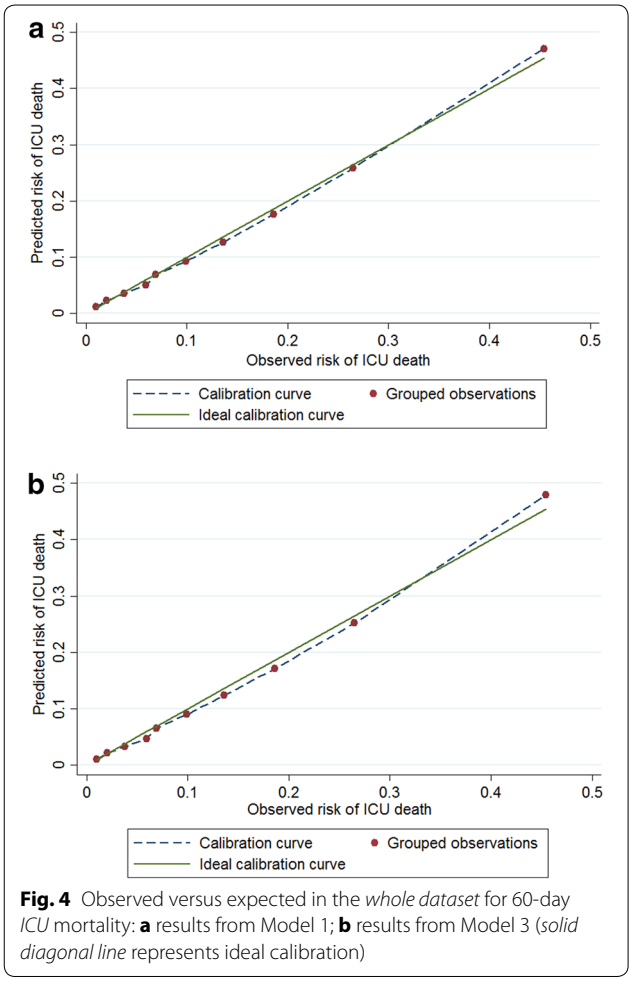


Fig. 4 Observed versus expected in the *whole dataset* for 60-day *ICU* mortality: **a** results from Model 1; **b** results from Model 3 (*solid diagonal line* represents ideal calibration)

Table 5 Exploratory analyses for model performance of Models 1 and 3 in different countries for 60-day hospital mortality using the whole dataset

Model performance	Goodness-of-fit		*C* index	SMR (95 % CI)
	p value[a]	AIC		
Model 1[b]				
Canada ($n = 2456$)	0.41	7639	0.61	0.981 (0.942–1.021)
Australia ($n = 768$)	0.11	1418	0.62	0.965 (0.897–1.037)
USA and UK ($n = 109$)	0.64	130	0.68	0.982 (0.804–1.186)
Saudi Arabia and Brazil ($n = 413$)	0.77	1713	0.55	1.155 (1.054–1.263)
Model 3[c]				
Canada ($n = 2456$)	0.88	7489	0.62	0.996 (0.956–1.037)
Australia ($n = 768$)	0.80	1349	0.63	0.968 (0.898–1.043)
USA and UK ($n = 109$)	0.41	125	0.71	0.981 (0.801–1.190)
Saudi Arabia and Brazil ($n = 413$)	0.25	1660	0.63	1.099 (0.943–1.311)

AIC Akaike information criterion, *SMR* standardized mortality ratio

[a] Based on Groennesby and Borgan test

[b] Model 1 included APACHE II score only

[c] Model 3 consisted of BMI, medical admission, inotropes or vasopressors, acetylsalicylic acid or clopidogrel, APACHE II score and the interaction between BMI and APACHE II score

Comparison with other studies

Prediction models based on multivariate analyses typically use logistic regression analysis, due to the advantage of its simpler interpretation of the relationship between predictive factors and outcomes [56]. One study built a prediction model combining APACHE II score, a Model for End-Stage Liver Disease score, mechanical ventilation and sex using logistic regression in ICU patients with end-stage liver disease [57]. They found that the new model was more accurate than APACHE II score alone (the area under the receiver operating characteristic curves (AUC): 0.86 versus 0.76) in prediction of hospital mortality [57]. Another cohort study employed an assessment tool based on the PIRO (predisposition, insult, response and organ dysfunction) concept including

Table 4 Sensitivity analyses of model performance in Model 3 using restricted cubic splines for continuous predictors[a] for 60-day hospital and 60-day ICU mortality

Model performance	Goodness-of-fit		*C* index	SMR (95 % CI)
	p value[b]	AIC		
Hospital mortality in derivation set ($n = 1891$)[c]	0.54	5691	0.63	1.007 (0.963–1.054)
Hospital mortality in validation set ($n = 1855$)[c]	0.51	5924	0.62	0.988 (0.943–1.034)
ICU mortality in the whole set ($n = 3746$)[d]	0.69	8177	0.64	1.005 (0.973–1.037)

AIC Akaike information criterion, *SMR* standardized mortality ratio

[a] Continuous predictors included BMI, APACHE score and the interaction between them for hospital mortality, and continuous predictors only included BMI and APACHE score for ICU mortality

[b] Based on Groennesby and Borgan test

[c] Model 3 for hospital mortality included BMI, medical admission, inotropes or vasopressors, acetylsalicylic acid or clopidogrel, APACHE II score and the interaction between BMI and APACHE II score

[d] Model 3 for ICU mortality consisted of BMI, medical admission, invasive mechanical ventilation and APACHE II score

comorbidities, old age, multilobar opacities in chest radiograph, shock, hypoxemia, bacteremia, acute renal failure and acute respiratory distress syndrome, to compare its model performance with APACHE II score in patients with community-acquired pneumonia [58]. The AUC of the PIRO score (0.88) was significantly higher than that of APACHE II score (0.75) in predicting 28-day ICU mortality [58]. Though it was difficult to directly compare our results with these models, given their different populations, settings, data and methodologies, these studies agreed with our findings in that adding more information to build a new model would likely outperform the APACHE II score alone.

Limitations and strengths

This study was based on the data from a randomized thromboprophylaxis trial with strict inclusion and exclusion criteria, which therefore limits the generalizability of its findings. For instance, the mortality rate in this study may be lower than in other studies, because patients with poor life expectancy were excluded in the trial protocol [34, 35]. As well, the population upon which the new model was developed excluded patients who were at high risk of bleeding or if they were admitted to ICU because of major trauma, neurosurgery or orthopedic surgery [34, 35]. The latter criteria could also explain the apparent lower mortality associated with surgical patients compared to medical patients, as some of the more seriously ill surgical patients (e.g., patients with trauma, neurosurgical or orthopedic surgery) were excluded from the study. In addition, we could only use the data included in the original trial database, and we could not subsequently capture other potentially important indicators of illness severity including those that might have helped with the discrimination of this new model.

Strengths of this study include the international multicenter design, large sample size and standardized data collection. Moreover, we performed rigorous statistical analyses to build a new model and evaluate its performance. Evidence from internal validation and sensitivity analyses indicated that the findings were internally validated and robust. Similar results from explanatory analyses in different countries also suggested the generalizability and robustness of the model in this dataset using a heterogeneous group of patients.

Conclusion

Using data from critically ill medical–surgical patients receiving heparin thromboprophylaxis, we identify additional risk factors for mortality independent of APACHE II score and construct a new model to predict risk of death. The new model combining APACHE II score and other risk factors is a good fit, well calibrated, but with unsatisfactory discriminative power. Compared with the APACHE II score alone, the new prediction model which increases data collection, is more complex but does not substantially improve discriminative ability.

Abbreviations

AIC: Akaike information criterion; APACHE: Acute Physiology and Chronic Health Evaluation; AUC: area under the receiver operating characteristic curves; BMI: body mass index; CI: confidence interval; HR: hazard ratio; IQR: interquartile range; ICU: intensive care unit; LMWH: low molecular weight heparin; PIRO: predisposition, insult, response and organ dysfunction; PROTECT: Prophylaxis for Thromboembolism in Critical Care Trial; RCT: randomized controlled trial; SMR: standardized mortality ratio; SD: standard deviation; TRIPOD: transparent reporting of a multivariable prediction model for individual prognosis or diagnosis; UFH: unfractionated heparin; VIF: variance inflation factor.

Authors' contributions

GL contributed to the study conception and design, data analysis, interpretation of the data and drafting the manuscript. LT, DC and ML contributed to the study design, interpretation of the data, critical revision of the manuscript and approval of the final version of the manuscript. RL and JM contributed to interpretation of the data, critical revision of the manuscript and approval of the final version of the manuscript. NA-D and DH-A contributed to the analysis and interpretation of the data, critical revision of the manuscript and approval of the final version of the manuscript. GG, AH, RF, NA, RT, YA, DC, PD, AF and SW contributed to critical version of the manuscript and approval of the final version of the manuscript. All authors read and approved the final manuscript.

Author details

[1] Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada. [2] Centre for Evaluation of Medicines, St. Joseph's Healthcare Hamilton, McMaster University, 25 Main St. West, Suite 2000, 20th Floor, Hamilton, ON L8P 1H1, Canada. [3] Department of Medicine, McMaster University, Hamilton, ON, Canada. [4] Duke Clinical Research Institute, Duke University, Durham, NC, USA. [5] Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, ON, Canada. [6] Critical Care Medicine, St. Michael's Hospital, Toronto, ON, Canada. [7] School of Nursing, McMaster University, Hamilton, ON, Canada. [8] Department of Critical Care Medicine, Sunnybrook Health Sciences Centre, Toronto, ON, Canada. [9] Mercy Clinic Adult Critical Care, Mercy Hospital Saint Louis, Saint Louis, MO, USA. [10] King Saud bin Abdulaziz University for Health Sciences and King Abdullah International Medical Research Center, Riyadh, Saudi Arabia. [11] Critical Care Medicine, Vancouver Coastal Health Research Institute, Vancouver, BC, Canada. [12] Center for Health Evaluation and Outcome Sciences and Division of Critical Care Medicine, St. Paul's Hospital and University of British Columbia, Vancouver, BC, Canada.

Acknowledgements

The authors thank the contributions of the bedside staff, research coordinators and all the patients in the trial herein.

Funding information

PROTECT was funded by the Canadian Institutes of Health Research (CIHR, #MCT78568), the Australian and New Zealand College of Anesthetists Research Foundation (#07/23) and the Heart and Stroke Foundation of Canada (#T6157, #T6950, #NA6186).

Competing interests

Mr. Li receives a Father Sean O'Sullivan Research Award, the Research Institute of St. Joseph's Healthcare Hamilton, and a doctoral award from the CSC. Dr. Cook is a research chair of the CIHR. Dr. Lopes receives research grant from Bristol-Myers Squibb and sits on advisory boards for Bristol-Myers Squibb, BI and Bayer. No potential conflicts of interest exist for the remaining authors.

Appendix

See Tables 6, 7, 8 and 9.

Table 6 Predictors for 30-day hospital mortality in the derivation dataset and for 30-day ICU mortality in the whole dataset

Predictors	Hospital mortality ($n = 1891$)[a]		ICU mortality ($n = 3746$)[b]	
	HR (95 % CI)	*p* value	HR (95 % CI)	*p* value
BMI	0.97 (0.95–0.99)	<0.001	0.97 (0.96–0.99)	<0.001
Medical admission	1.86 (1.38–2.51)	<0.001	1.41 (1.12–1.78)	0.003
Inotropes or vasopressors	1.45 (1.16–1.82)	0.001	–[c]	–[c]
Acetylsalicylic acid or clopidogrel	1.22 (1.02–1.46)	0.037	–[c]	–[c]
APACHE II score	1.03 (1.01–1.04)	<0.001	1.04 (1.03–1.05)	<0.001
Invasive mechanical ventilation	–[d]	–[d]	0.77 (0.59–0.99)	0.042

[a] There were 332 deaths for 30-day hospital mortality in derivation cohort

[b] There were 522 deaths for 30-day ICU mortality in the whole cohort

[c] Not included in the model for ICU mortality

[d] Not included in the model for hospital mortality

Table 7 Results for sensitivity analyses of model performance in the new model for 30-day hospital mortality and 30-day ICU mortality

Model performance	Goodness-of-fit		*C* index	SMR (95 % CI)
	p value[a]	AIC		
Hospital mortality in derivation set ($n = 1891$)[b]	0.74	4682	0.63	1.005 (0.960–1.051)
Hospital mortality in validation set ($n = 1855$)[b]	0.76	4789	0.65	1.006 (0.961–1.053)
ICU mortality in the whole set ($n = 3746$)[c]	0.65	7559	0.64	1.003 (0.971–1.036)

AIC Akaike information criterion, *SMR* standardized mortality ratio

[a] Based on Groennesby and Borgan test

[b] The new model for 30-day hospital mortality included BMI, medical admission, inotropes or vasopressors, acetylsalicylic acid or clopidogrel and APACHE II score

[c] The new model for 30-day ICU mortality consisted of BMI, medical admission, invasive mechanical ventilation and APACHE II score

Table 8 Predictors for 90-day hospital mortality in the derivation dataset and for 90-day ICU mortality in the whole dataset

Predictors	Hospital mortality ($n = 1891$)[a]		ICU mortality ($n = 3746$)[b]	
	HR (95 % CI)	*p* value	HR (95 % CI)	*p* value
BMI	0.93 (0.88–0.97)	0.003	0.98 (0.96–0.99)	<0.001
Medical admission	1.67 (1.29–2.17)	<0.001	1.39 (1.12–1.73)	0.003
Inotropes or vasopressors	1.36 (1.11–1.67)	0.003	–[c]	–[c]
Acetylsalicylic acid or clopidogrel	1.27 (1.02–1.59)	0.035	–[c]	–[c]
APACHE II score	0.97 (0.92–1.02)	0.241	1.04 (1.02–1.05)	<0.001
APACHE II score*BMI	1.002 (1.000–1.004)	0.038	–[c]	–[c]
Invasive mechanical ventilation	–[d]	–[d]	0.75 (0.58–0.97)	0.026

[a] There were 405 deaths for 90-day hospital mortality in derivation cohort

[b] There were 581 deaths for 90-day ICU mortality in the whole cohort

[c] Not included in the model for ICU mortality

[d] Not included in the model for hospital mortality

Table 9 Results for sensitivity analyses of model performance in the new model for 90-day hospital mortality and 90-day ICU mortality

Model performance	Goodness-of-fit		*C* index	SMR (95 % CI)
	p value[a]	AIC		
Hospital mortality in derivation set ($n = 1891$)[b]	0.40	5521	0.63	1.006 (0.962–1.053)
Hospital mortality in validation set ($n = 1855$)[b]	0.62	5787	0.64	1.008 (0.963–1.055)
ICU mortality in the whole set ($n = 3746$)[c]	0.78	8144	0.63	1.004 (0.972–1.037)

AIC Akaike information criterion, *SMR* standardized mortality ratio

[a] Based on Groennesby and Borgan test

[b] The new model for 90-day hospital mortality included BMI, medical admission, inotropes or vasopressors, acetylsalicylic acid or clopidogrel, APACHE II score and the interaction between BMI and APACHE II score

[c] The new model for 90-day ICU mortality consisted of BMI, medical admission, invasive mechanical ventilation and APACHE II score

Received: 5 November 2015 Accepted: 2 February 2016
Published online: 27 February 2016

References

1. Checkley W, Martin GS, Brown SM, Chang SY, Dabbagh O, Fremont RD, Girard TD, Rice TW, Howell MD, Johnson SB, et al. Structure, process, and annual ICU mortality across 69 centers: United States Critical Illness and Injury Trials Group Critical Illness Outcomes Study. Crit Care Med. 2014;42(2):344–56.
2. Wu AW, Pronovost P, Morlock L. ICU incident reporting systems. J Crit Care. 2002;17(2):86–94.
3. Parker J. California Intensive Care Outcomes Project (CALICO). http://www.oshpd.ca.gov/HID/Products/PatDischargeData/ICUDataCALICO/CALICO_05-07.pdf (2007). Accessed 20 Feb 2015.
4. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. Contin Educ Anaesth Critical Care Pain. 2008;8(5):181–5.
5. Gartman EJ, Casserly BP, Martin D, Ward NS. Using serial severity scores to predict death in ICU patients: a validation study and review of the literature. Curr Opin Crit Care. 2009;15(6):578–82.
6. Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, de Jonge E, Bosman RJ, Peelen L, de Keizer NF. External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. J Crit Care. 2011;26(1):105.e11–8.
7. Zimmerman JE, Kramer AA. Outcome prediction in critical care: the Acute Physiology and Chronic Health Evaluation models. Curr Opin Crit Care. 2008;14(5):491–7.
8. Beck DH, Smith GB, Pappachan JV, Millar B. External validation of the SAPS II, APACHE II and APACHE III prognostic models in South England: a multicentre study. Intensive Care Med. 2003;29(2):249–56.
9. Wong DT, Knaus WA. Predicting outcome in critical care: the current status of the APACHE prognostic scoring system. Can J Anaesth. 1991;38(3):374–83.
10. Oh TE, Hutchinson R, Short S. BUCKLEY T, LIN E, LEUNG D: Verification of the acute physiology and chronic health evaluation scoring system in a Hong Kong intensive care unit. Crit Care Med. 1993;21(5):698–705.
11. Capuzzo M, Valpondi V, Sgarbi A, Bortolazzi S, Pavoni V, Gilli G, Candini G, Gritti G, Alvisi R. Validation of severity scoring systems SAPS II and APACHE II in a single-center population. Intensive Care Med. 2000;26(12):1779–85.
12. Knaus WA, Draper EA, Wagner DP, Zimmerman JE, Birnbaum ML, Cullen DJ, Kohles MK, Shin B, Snyder JV. Evaluating outcome from intensive care: a preliminary multihospital comparison. Crit Care Med. 1982;10(8):491–6.
13. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–29.
14. Knaus WA, Wagner D, Draper E, Zimmerman J, Bergner M, Bastos PG, Sirio C, Murphy D, Lotring T, Damiano A. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest. 1991;100(6):1619–36.
15. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients*. Crit Care Med. 2006;34(5):1297–310.
16. Arabi Y, Al Shirawi N, Memish Z, Venkatesh S, Al-Shimemeri A. Assessment of six mortality prediction models in patients admitted with severe sepsis and septic shock to the intensive care unit: a prospective cohort study. Crit Care. 2003;7(5):R116.
17. Keegan MT, Gali B, Findlay JY, Heimbach JK, Plevak DJ, Afessa B. APACHE III outcome prediction in patients admitted to the intensive care unit after liver transplantation: a retrospective cohort study. BMC Surg. 2009;9:11.
18. Namendys-Silva SA, Baltazar-Torres JA, Rivero-Sigarroa E, Fonseca-Lazcano JA, Montiel-Lopez L, Dominguez-Cherit G. Prognostic factors in patients with systemic lupus erythematosus admitted to the intensive care unit. Lupus. 2009;18(14):1252–8.
19. Papachristou GI, Muddana V, Yadav D, O'Connell M, Sanders MK, Slivka A, Whitcomb DC. Comparison of BISAP, Ranson's, APACHE-II, and CTSI scores in predicting organ failure, complications, and mortality in acute pancreatitis. Am J Gastroenterol. 2010;105(2):435–41.
20. Inal MT, Memis D, Kargi M, Sut N. Prognostic value of indocyanine green elimination assessed with LiMON in septic patients. J Crit Care. 2009;24(3):329–34.
21. Siontis G, Tzoulaki I, Ioannidis J. Predicting death: an empirical evaluation of predictive tools for mortality. Arch Intern Med. 2011;171(19):1721–6.
22. Lagu T, Rothberg MB, Nathanson BH, Steingrub JS, Lindenauer PK. Incorporating initial treatments improves performance of a mortality prediction model for patients with sepsis. Pharmacoepidemiol Drug Saf. 2012;21(S2):44–52.
23. O'Brien JM Jr, Phillips GS, Ali NA, Lucarelli M, Marsh CB, Lemeshow S. Body mass index is independently associated with hospital mortality in mechanically ventilated adults with acute lung injury. Crit Care Med. 2006;34(3):738.
24. Pickkers P, de Keizer N, Dusseljee J, Weerheijm D, van der Hoeven JG, Peek N. Body mass index is associated with hospital mortality in critically ill patients: an observational cohort study. Crit Care Med. 2013;41(8):1878–83.
25. Mahmood K, Eldeirawi K, Wahidi MM. Association of gender with outcomes in critically ill patients. Crit Care. 2012;16(3):R92.
26. Combes A, Luyt CE, Trouillet JL, Nieszkowska A, Chastre J. Gender impact on the outcomes of critically ill patients with nosocomial infections. Crit Care Med. 2009;37(9):2506–11.
27. Nachtigall I, Tafelski S, Rothbart A, Kaufner L, Schmidt M, Tamarkin A, Kartachov M, Zebedies D, Trefzer T, Wernecke KD, et al. Gender-related outcome difference is related to course of sepsis on mixed ICUs: a prospective, observational clinical study. Crit Care. 2011;15(3):R151.
28. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR. SAPS 3–From evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. 2005;31(10):1345–55.
29. Walsh TS, Stanworth SJ, Prescott RJ, Lee RJ, Watson DM, Wyncoll D. Prevalence, management, and outcomes of critically ill patients with

prothrombin time prolongation in United Kingdom intensive care units. Crit Care Med. 2010;38(10):1939–46.
30. Angstwurm MW, Dempfle CE, Spannagl M. New disseminated intravascular coagulation score: a useful tool to predict mortality in comparison with Acute Physiology and Chronic Health Evaluation II and Logistic Organ Dysfunction scores. Crit Care Med. 2006;34(2):314–20.
31. Williamson DR, Albert M, Heels-Ansdell D, Arnold DM, Lauzier F, Zarychanski R, Crowther M, Warkentin TE, Dodek P, Cade J, et al. Thrombocytopenia in critically ill patients receiving thromboprophylaxis: frequency, risk factors, and outcomes. Chest. 2013;144(4):1207–15.
32. Strauss R, Wehler M, Mehler K, Kreutzer D, Koebnick C, Hahn EG. Thrombocytopenia in patients in the medical intensive care unit: bleeding prevalence, transfusion requirements, and outcome. Crit Care Med. 2002;30(8):1765–71.
33. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
34. Cook D, Meade M, Guyatt G, Walter SD, Heels-Ansdell D, Geerts W, Warkentin TE, Cooper DJ, Zytaruk N, Vallance S. PROphylaxis for ThromboEmbolism in Critical Care Trial protocol and analysis plan. J Crit Care. 2011;26(2):e221–3.
35. Cook D, Meade M, Guyatt G, Walter S, Heels-Ansdell D, Warkentin TE, Zytaruk N, Crowther M, Geerts W, Cooper DJ, et al. Dalteparin versus unfractionated heparin in critically ill patients. New Engl J Med. 2011;364(14):1305–14.
36. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. Am Stat. 2007;61(1):79–90.
37. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ. 2009;338:b604. doi:10.1136/bmj.b604.
38. Miles J, Shevlin M. Appling regression and correlation: a guide for students. London: Sage; 2001.
39. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515–26.
40. Goldhill DR, Sumner A. Outcome of intensive care patients in a group of British intensive care units. Crit Care Med. 1998;26(8):1337–45.
41. Nijman RG, Vergouwe Y, Thompson M, van Veen M, van Meurs AH, van der Lei J, Steyerberg EW, Moll HA, Oostenbrink R: Clinical prediction model to aid emergency doctors managing febrile children at risk of serious bacterial infections: diagnostic study. BMJ. 2013;346:f1706.
42. Gronnesby JK, Borgan O. A method for checking regression models in survival analysis based on the risk score. Lifetime Data Anal. 1996;2(4):315–28.
43. Akaike H. A new look at the statistical model identification. Autom Control IEEE Trans. 1974;19(6):716–23.
44. Harrell F, Lee KL, Mark DB. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15:361–87.
45. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.
46. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013;13(1):33.
47. Pickkers P, de Keizer N, Dusseljee J, Weerheijm D, van der Hoeven JG, Peek N. Body mass index is associated with hospital mortality in critically ill patients: an observational cohort study. Crit Care Med. 2013;41(8):1878–83.
48. Hogue CW Jr, Stearns JD, Colantuoni E, Robinson KA, Stierer T, Mitter N, Pronovost PJ, Needham DM. The impact of obesity on outcomes after critical illness: a meta-analysis. Intensive Care Med. 2009;35(7):1152–70.
49. Hutagalung R, Marques J, Kobylka K, Zeidan M, Kabisch B, Brunkhorst F, Reinhart K, Sakr Y. The obesity paradox in surgical intensive care unit patients. Intensive Care Med. 2011;37(11):1793–9.
50. Kalantar-Zadeh K, Block G, Horwich T, Fonarow GC. Reverse epidemiology of conventional cardiovascular risk factors in patients with chronic heart failure. J Am Coll Cardiol. 2004;43(8):1439–44.
51. Niedziela J, Hudzik B, Niedziela N, Gasior M, Gierlotka M, Wasilewski J, Myrda K, Lekston A, Polonski L, Rozentryt P. The obesity paradox in acute coronary syndrome: a meta-analysis. Eur J Epidemiol. 2014;29(11):801–12.
52. Lasocki S. The true obesity paradox: obese and malnourished? Crit Care Med. 2015;43(1):240–1.
53. Oliveros H, Villamor E. Obesity and mortality in critically ill adults: a systematic review and meta-analysis. Obesity. 2008;16(3):515–21.
54. Molina R, Bernal T, Borges M, Zaragoza R, Bonastre J, Granada RM, Rodriguez-Borregán JC, Núñez K, Seijas I, Ayestaran I. Ventilatory support in critically ill hematology patients with respiratory failure. Crit Care. 2012;16(4):R133.
55. van Gestel J, Bierings M, Dauger S, Dalle J, Pavlíček P, Sedláček P, Monteiro L, Lankester A, Bollen C. Outcome of invasive mechanical ventilation after pediatric allogeneic hematopoietic SCT: results from a prospective, multicenter registry. Bone Marrow Transplant. 2014;49(10):1287–92.
56. Adams ST, Leveson SH: Clinical prediction rules. BMJ. 2012;344:d8312.
57. Balekian AA, Gould MK. Predicting in-hospital mortality among critically ill patients with end-stage liver disease. J Crit Care. 2012;27(6):e741–7.
58. Rello J, Rodriguez A, Lisboa T, Gallego M, Lujan M, Wunderink R. PIRO score for community-acquired pneumonia: a new prediction rule for assessment of severity in intensive care unit patients with community-acquired pneumonia*. Crit Care Med. 2009;37(2):456–62.

Chapter 5

Competing Risk Analysis for Evaluation of Dalteparin Versus Unfractionated Heparin for Venous Thromboembolism in Medical-Surgical Critically Ill Patients

Guowei Li, MBBS, MSc, Deborah J. Cook, MD, Mitchell A.H. Levine, MD, Gordon Guyatt, MD, Mark Crowther, MD, Diane Heels-Ansdell, MSc, Anne Holbrook, MD, PharmD, Francois Lamontagne, MD, Stephen D. Walter, PhD, Niall D. Ferguson, MD, Simon Finfer, MD, Yaseen M. Arabi, MD, Rinaldo Bellomo, MD, and D. Jamie Cooper, MD, PhD, Lehana Thabane, PhD, the PROTECT Investigators for the Canadian Critical Care Trials Group, and the Australian and New Zealand Intensive Care Society Clinical Trials Group

Abstract: Failure to recognize the presence of competing risk or to account for it may result in misleading conclusions. We aimed to perform a competing risk analysis to assess the efficacy of the low molecular weight heparin dalteparin versus unfractionated heparin (UFH) in venous thromboembolism (VTE) in medical-surgical critically ill patients, taking death as a competing risk.

This was a secondary analysis of a prospective randomized study of the Prophylaxis for Thromboembolism in Critical Care Trial (PROTECT) database. A total of 3746 medical-surgical critically ill patients from 67 intensive care units (ICUs) in 6 countries receiving either subcutaneous UFH 5000 IU twice daily (n = 1873) or dalteparin 5000 IU once daily plus once-daily placebo (n = 1873) were included for analysis.

A total of 205 incident proximal leg deep vein thromboses (PLDVT) were reported during follow-up, among which 96 were in the dalteparin group and 109 were in the UFH group. No significant treatment effect of dalteparin on PLDVT compared with UFH was observed in either the competing risk analysis or standard survival analysis (also known as cause-specific analysis) using multivariable models adjusted for APACHE II score, history of VTE, need for vasopressors, and end-stage renal disease: sub-hazard ratio (SHR) = 0.92, 95% confidence interval (CI): 0.70–1.21, *P*-value = 0.56 for the competing risk analysis; hazard ratio (HR) = 0.92, 95% CI: 0.68–1.23, *P*-value = 0.57 for cause-specific analysis. Dalteparin was associated with a significant reduction in risk of pulmonary embolism (PE): SHR = 0.54, 95% CI: 0.31–0.94, *P*-value = 0.02 for the competing risk analysis; HR = 0.51, 95% CI: 0.30–0.88, *P*-value = 0.01 for the cause-specific analysis. Two additional sensitivity analyses using the treatment variable as a time-dependent covariate and using as-treated and per-protocol approaches demonstrated similar findings.

This competing risk analysis yields no significant treatment effect on PLDVT but a superior effect of dalteparin on PE compared with UFH in medical-surgical critically ill patients. The findings from the competing risk method are in accordance with results from the cause-specific analysis.

clinicaltrials.gov Identifier: NCT00182143

(*Medicine* 94(36):e1479)

Abbreviations: APACHE = Acute Physiology and Chronic Health Evaluation, CI = confidence interval, CIF = cumulative incidence function, DVT = deep vein thrombosis, HR = hazard ratio, ICU = intensive care unit, IQR = interquartile range, PE = pulmonary embolism, PLDVT = proximal leg deep vein thromboses, PROTECT = Prophylaxis for Thromboembolism in Critical Care Trial, RCT = randomized controlled trial, SD = standard deviation, SHR = sub-hazard ratio, UFH = unfractionated heparin, VTE = venous thromboembolism.

INTRODUCTION

A competing risk is defined as an event that either precludes another event under investigation or fundamentally alters

Editor: Miao Liu.
Received: June 5, 2015; revised: August 6, 2015; accepted: August 7, 2015.
From the Department of Clinical Epidemiology and Biostatistics (GL, DC, MAHL, GG, MC, DHA, AH, SDW, LT); Department of Medicine (DC, MAHL, GG, AH); St. Joseph's Healthcare Hamilton, McMaster University, Hamilton, ON (DJC, MAHL, MC, AH, LT); Centre de recherche du CHU de Sherbrooke et Université de Sherbrooke, Sherbrooke, QC (FL); Interdepartmental Division of Critical Care Medicine, Departments of Medicine and Physiology, and Institute for Health Policy, Management and Evaluation, University of Toronto (NDF); Department of Medicine, Division of Respirology, University Health Network and Mount Sinai Hospital (NDF); Toronto General Research Institute, Toronto, ON, Canada (NDF); Critical Care and Trauma Division, The George Institute, Sydney, Australia (SF); King Saud bin Abdulaziz University for Health Sciences and King Abdullah International Medical Research Center, Riyadh, Saudi Arabia (YMA); Australian and New Zealand Intensive Care Research Centre (RB, DC); School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia (DC).

Correspondence: Lehana Thabane, Clinical Epidemiology and Biostatistics, McMaster University, Father Sean O'Sullivan Research Centre, St. Joseph's Healthcare Hamilton, 3rd floor Martha Wing, 50 Charlton Avenue East, Hamilton, ON L8N 4A6, Canada (e-mail: thabanl@mcmaster.ca).

Supplemental Digital Content is available for this article.

PROTECT was funded by the Canadian Institutes of Health Research (CIHR, #MCT78568), the Australian and New Zealand College of Anesthetists Research Foundation (#07/23), and the Heart and Stroke Foundation of Canada (#T6157, #T6950, #NA6186). Mr Li receives a Father Sean O'Sullivan Research Award, the Research Institute of St. Joe's Hamilton, and a doctoral award from the China Scholarship Council. Dr Cook is a research chair of the CIHR. Dr Crowther holds a Career Investigator Award from the Heart and Stroke Foundation of Ontario and the Leo Pharma Chair in Thromboembolism Research at McMaster University and St Joseph's Healthcare, Hamilton. Dr Crowther sits on advisory boards for Leo Pharma Inc; Pfizer, Inc; Bayer; Boehringer-Ingelheim GmbH; Alexion; CSL Behring; and Artisan Pharma, Inc. He also prepares educational materials for Pfizer, Octapharma, and CSL Behring, and provides expert testimony for Bayer. Dr Lamontagne is a recipient of a research career award from the FRQ-S and a CIHR RCT Mentoring Award.

The authors have no conflicts of interest to disclose.

ISSN: 0025-7974
DOI: 10.1097/MD.0000000000001479

the probability of the outcome of interest.[1,2] In health research, it is not uncommon for participants to experience a competing risk event such as death, which prevents observing the event of interest. Failure to recognize the presence of competing risk or to account for it may result in misleading conclusions in clinical trials or epidemiological research.[3]

Critically ill patients in intensive care units (ICUs) are at high risk of venous thromboembolism (VTE), including deep vein thrombosis (DVT) and pulmonary embolism (PE), because of their complex acute and chronic illnesses, analgesia and paralysis, immobility, and other interventions they may receive.[4–6] Until recently, there were insufficient data to adequately compare the efficacy of low-molecular-weight heparin and unfractionated heparin (UFH) in preventing VTE in medical-surgical critically ill patients.[7] The multicenter international randomized controlled trial, PROTECT (Prophylaxis for Thromboembolism in Critical Care Trial), evaluated the efficacy of dalteparin (a low-molecular-weight heparin) versus UFH in proximal leg deep vein thromboses (PLDVT), and other VTEs.[8] The trial reported no significant effect of dalteparin versus UFH on PLDVT, but a significantly superior treatment effect of dalteparin on PE using standard survival analysis (also known as cause-specific analysis).[8] However, the mortality rate (23.3%) was much higher than the rate of PLDVT (5.5%) and PE (1.8%) during follow-up.

Death prior to a VTE precludes the occurrence of subsequent PLDVT and PE, and therefore it can potentially affect the estimation of thromboprophylaxis efficacy. Evidence has shown that cause-specific analyses that fail to take competing risks into account could report biased findings about the effect of treatments or prognostic factors on outcomes.[3,9,10] Cox regression used for cause-specific analyses may not be appropriate since its assumptions of noninformative censoring and independence of time distributions between PLDVT and death may have been violated because of the existence of competing risks.[2,11] Although in the original trial report, a composite outcome of VTE or death was used to examine the efficacy of dalteparin versus UFH,[8,12] death as a competing event for VTE was not directly accounted for in the analysis.

In this study, we reanalyzed data from PROTECT to explicitly account for death as a competing risk. We used the Fine and Gray proportional subdistribution hazards model that, emerging evidence suggests, is appropriate to use in the presence of competing risk[13,14] to evaluate the efficacy of dalteparin versus UFH in preventing VTE in medical-surgical critically ill patients. Our goal was to perform the competing risk analysis as a sensitivity analysis,[15] and thus assess the robustness of the main findings based on the cause-specific analysis.[8] We performed additional sensitivity analyses: using the treatment variable as a time-dependent covariate in both cause-specific analysis and the Fine and Gray model; and using as-treated and per-protocol approaches. Our primary outcome was PLDVT, and the secondary outcome was PE.

METHODS

Patients and Settings

Details about the design, conduct, and main results of PROTECT (ClinicalTrials.gov Identifier: NCT00182143) have been published elsewhere.[8,12] Briefly, PROTECT was a multicenter randomized controlled trial (RCT) conducted in 67 ICUs from 2006 to 2010 in Canada, the United States, Australia, Brazil, Saudi Arabia, and the United Kingdom, aiming to evaluate the efficacy of subcutaneous UFH 5000 IU twice daily versus dalteparin 5000 IU once daily plus once-daily placebo in VTE in 3746 medical-surgical critically ill patients. Patients were enrolled in this trial if they were $\geq$18-years old, weighed $\geq$45 kg, and were expected to stay in the ICU for $\geq$3 days. Exclusion criteria were an admission diagnosis of trauma, orthopedic surgery, uncontrolled hypertension, or neurosurgery; major hemorrhage within the previous week; stroke, coagulopathy, or thrombocytopenia; pregnancy; or limitation of life-support. Patients with a need for anticoagulant therapy, with a contraindication to heparin or blood products, or who were already enrolled in a related trial, were also excluded. All patients or their surrogates provided written informed consent. Research ethics committees at each center approved the trial (e-Table 1, http://links.lww.com/MD/A398).

Outcomes

The primary outcome was incident PLDVT detected $\geq$3 days postrandomization using bilateral proximal leg venous ultrasounds.[8,12] The screening ultrasonography was performed twice-weekly and if PLDVT was clinically suspected. The secondary outcome was incident PE. Pulmonary emboli were diagnosed when intraluminal filling defects appeared on computed tomography, or when an unmatched perfusion defect on ventilation–perfusion (V/Q) scans existed, or if there were both a pretest probability (clinical suspicion) and a nondiagnostic result on noninvasive testing.[8,12]

Two adjudicators were randomly assigned to independently assess the PLDVT events; 4 adjudicators evaluated each PE event. All adjudicators were blinded to treatment allocation, center, and each other's assessments.[8,12] All enrolled patients were followed up to hospital discharge to record their vital status. Data were censored at 100 days for VTE outcomes.[8,12]

Statistical Analyses

Descriptive statistics for baseline characteristics of the patients were presented as means and standard deviations (SDs) or median and interquartile range (IQR) for data on continuous variables, and frequencies (percentages) for categorical variables. Because the percentage of missing data was small ($<$5%),[8] we imputed the missing data using the mean or median of the variable in its group when survival analysis was performed. All tests were 2-sided at a significance level of 0.05.

The independence assumption of the time distribution between VTE and death may not be satisfied in survival analysis due to the competing risk of death, therefore the Kaplan–Meier method was not appropriate to estimate survival curves for VTE.[2,16] We used the cumulative incidence function (CIF, also known as the subdistribution), which was derived from the cause-specific hazard function and did not require the independence assumption, to estimate the marginal probability of VTE in the presence of competing risk.[13] Specifically, given a time point t, the CIF denoted the probability of experiencing a VTE by the time t when the patients could also die before they developed a VTE. We used the Pepe and Mori method to test whether the CIFs of VTE between the treatment groups (dalteparin versus UFH) were significantly different.[17] The Kaplan–Meier method was used to evaluate the CIF of death, and the log-rank test was performed to compare the CIFs between the treatment groups.[16]

All the analyses were conducted based on the intention-to-treat principle. We first performed univariate analyses for PLDVT and PE in the Fine and Gray model, and Cox regression for cause-specific analysis, respectively. Multivariable analyses

TABLE 1. Baseline Characteristics of the Dalteparin and UFH Group

Characteristics	Dalteparin Group*(n = 1873)	UFH Group† (n = 1873)
Age (year): mean (SD)	61.1 (16.5)	61.7 (16.4)
Gender: n, %		
Male	1052 (56.4)	1061 (57.0)
Female	813 (43.6)	801 (43.0)
APACHE II score: mean (SD)	21.4 (7.8)	21.7 (7.8)
History of personal or family VTE: n, %	86 (4.6)	87 (4.7)
Need for vasopressors: n, %	805 (43.2)	872 (46.8)
End-stage renal failure: n, %	60 (3.2)	58 (3.1)

APACHE = Acute Physiology and Chronic Health Evaluation; SD = standard deviation; UFH = unfractionated heparin; VTE = venous thromboembolism.
* Median follow-up: 18 days; interquartile range: 10–32 days.
† Median follow-up: 17 days; interquartile range: 10–35 days.

were then employed, in which the analyses were adjusted for the Acute Physiology and Chronic Health Evaluation (APACHE) II score, history of personal or family VTE, need for vasopressors, and end-stage renal disease.[8,18] Sub-hazard ratios (SHRs) and corresponding 95% confidence intervals (CIs) were reported for the Fine and Gray model, while hazard ratios (HRs) were presented for cause-specific analysis. Both a statistical test and a graphical examination based on the Schoenfeld residuals were used to assess the proportional hazards assumption.[13,19]

Two additional sensitivity analyses were conducted. Since there may be gaps in the treatment of participants during follow-up, we included the treatment as a time-dependent covariate in both cause-specific analysis and the Fine and Gray model, to investigate whether the estimated treatment effect was robust.[20] Another sensitivity analysis was performed by using as-treated and per-protocol analyses in PLDVT and PE. The as-treated analysis excluded the patients (n = 87) who withdrew consent, never received any study drug, or who were incorrectly randomized.[8,12] The per-protocol analysis excluded the patients (n = 619) who were treated for baseline VTE diagnosed on the first screening ultrasonography, had <2 ultrasound tests, or who received study treatment for <2 days.[8,12]

All analyses were performed using STATA Version 12 (Stata Corp., College Station, TX) and SAS version 9.3 (SAS Institute, Inc., Cary, NC).

RESULTS

Baseline Characteristics

The selection process of patients in the PROTECT has been published elsewhere.[8] Briefly, in the intension-to-treat analysis, 3746 patients (43.3% females) were included. Their mean age was 61.4 (SD: 16.5) years, and their mean APACHE II score was 21.5 (SD: 7.76) at baseline. The 4.6% of the patients (n = 173) had a history of personal or family VTE, and the percentage of participants diagnosed as end-stage renal disease was 3.2% (n = 118). There were 1677 (44.8%) patients requiring vasopressors at baseline.

The baseline characteristics of the dalteparin and UFH groups are shown in Table 1. There were 1873 participants assigned to dalteparin group and 1873 patients to UFH, respectively. The age, sex composition, APACHE II scores, the percentages having a history of VTE or a diagnosis of end-stage renal disease, and the numbers of patients requiring vasopressors were similar in the 2 groups. The median follow-up for the dalteparin group was 18 days (IQR: 10–32), while the median follow-up for the UFH group was 17 days (IQR: 10–35) (Table 1).

During follow-up, 205 incident PLDVTs were reported, among which 96 were in the dalteparin group and 109 were in the UFH group. The 96 patients with PLDVT in the dalteparin group had similar age (60.2 versus 61.8), female composition (51.04% versus 43.12%), APACHE II scores (22.6 versus 22.4), percentage of history of VTE (6.25% versus 8.26%), diagnosis of end-stage renal disease (5.21% versus 2.75%), and patients requiring vasopressors (47.92% versus 55.96%) to the 109 patients in the UFH group (all P-values > 0.20). There were 812 patients (386 and 426 in the dalteparin and UFH group, respectively) who died before they developed a PLDVT during their ICU and hospital stay. The 30-, 60-, and 90-day cumulative incidence of PLDVT for dalteparin compared with UFH group was 11.8% versus 12.2%, 16.2% versus 15.7%, and 17.7% versus 15.7%, respectively. The cumulative incidence curves of PLDVT for these 2 groups are displayed in Figure 1 with considerable overlap. No significant difference of the CIFs for PLDVT between the 2 groups was observed using the Pepe

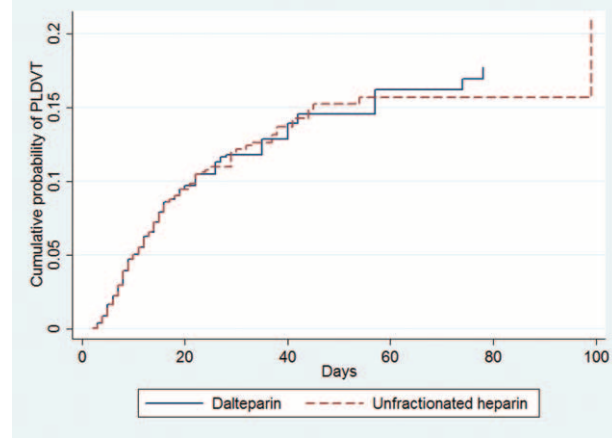


FIGURE 1. Cumulative incidence curves of PLDVT in dalteparin and UFH group. PLDVT = proximal leg deep vein thrombosis, UFH = unfractionated heparin.

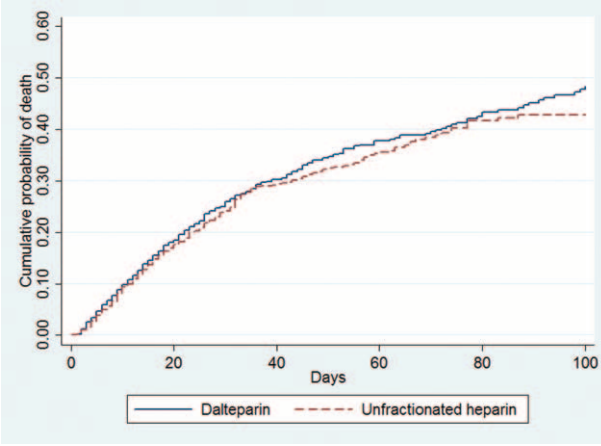


FIGURE 2. Cumulative incidence curves of death in dalteparin and unfractionated heparin (UFH) group using the Kaplan–Meier method.

and Mori test (P-value = 0.66). The cumulative incidence curves of mortality using the Kaplan–Meier method between the dalteparin group and UFH group are shown in Figure 2. Similarly no significant difference of these 2 CIFs for death was found (P-value = 0.23 for log-rank test).

Comparison Between Competing Risk Analysis and Cause-Specific Analysis

Results from the Fine and Gray model and the cause-specific method to evaluate the efficacy of dalteparin versus UFH in PLDVT are presented in Table 2. No significant treatment effect of dalteparin on PLDVT was observed in either the competing risk analysis or the cause-specific analysis using univariate analyses: SHR = 0.92, 95% CI: 0.71–1.21, P-value = 0.56 for the Fine and Gray model; HR = 0.92, 95% CI: 0.70–1.21, P-value = 0.54 for the cause-specific method. Similar findings were also identified in multivariable models adjusted for APACHE II score, history of VTE, need for vasopressors, and end-stage renal disease: SHR = 0.92, 95% CI: 0.70–1.21, P-value = 0.56 for the Fine and Gray model; HR = 0.92, 95% CI: 0.68–1.23, P-value = 0.57 for the cause-specific analysis (Table 2). Moreover, we performed another multivariable analysis in both competing risk analysis and cause-specific analysis adjusted for age, female gender, APACHE II score, history of VTE, need for vasopressors, and end-stage renal disease. Findings remained consistent: SHR = 0.92, 95% CI: 0.70–1.22, P-value = 0.55 for the Fine and Gray model; HR = 0.92, 95% CI: 0.69–1.21, P-value = 0.54 for the cause-specific analysis.

Table 2 also shows results for evaluation of dalteparin versus UFH for the outcome of PE. There were 24 patients with incident PE reported in the dalteparin group, while 43 patients were diagnosed with PE in the UFH group. Dalteparin was associated with significantly fewer PE compared with UFH in the univariate analysis, with a SHR of 0.58 (P-value = 0.03) for the competing risk analysis and a HR of 0.58 (P-value = 0.03) for the cause-specific method. The significant treatment effect of dalteparin remained unchanged in the multivariable analysis in both the Fine and Gray model (SHR = 0.54, 95% CI: 0.31–0.94, P-value = 0.02) and the cause-specific analysis (HR = 0.51, 95% CI: 0.30–0.88, P-value = 0.01), compared with UFH (Table 2).

Additional Sensitivity Analyses

Table 3 shows results of additional sensitivity analyses including the treatment as a time-dependent covariate in both cause-specific analysis and the Fine and Gray model. Similar findings to the analyses using dalteparin and UFH as time-invariant covariates (Table 2) were reported: no significant treatment effect of dalteparin on PLDVT compared with UFH was found (P-values $\geq$ 0.50), while a significantly protective effect on PE was observed (P-values $<$ 0.05) in both the competing risk analysis and the cause-specific method (Table 3).

Another sensitivity analysis was conducted by using as-treated and per-protocol multivariable analyses in PLDVT and

TABLE 2. Univariate and Multivariable Analyses in PLDVT and PE in the Fine and Gray Model and the Cause-Specific Method

	PLDVT				PE			
Method	**Dalteparin (n = 1873)**	**UFH (n = 1873)**	**Statistics* (95% CI)**	***P*-Value**	**Dalteparin (n = 1873)**	**UFH (n = 1873)**	**Statistics* (95% CI)**	***P*-Value**
Univariate analysis								
Fine and Gray model	96 (5.13)†	109 (5.82)†	0.92 (0.71–1.21)	0.56	24 (1.28)†	43 (2.30)†	0.58 (0.35–0.95)	0.03
Cause-specific analysis			0.92 (0.70–1.21)	0.54			0.58 (0.34–0.96)	0.03
Multivariable analysis‡								
Fine and Gray model	96 (5.13)†	109 (5.82) †	0.92 (0.70–1.21)	0.56	24 (1.28)†	43 (2.30)†	0.54 (0.31–0.94)	0.02
Cause-specific analysis			0.92 (0.68–1.23)	0.57			0.51 (0.30–0.88)	0.01

PE = pulmonary embolism, PLDVT = proximal leg deep vein thrombosis, UFH = unfractionated heparin.
* Sub-hazard ratio (SHR) for the Fine and Gray model; hazard ratio (HR) for the cause-specific analysis.
† Expressed as the number and percentage (%) of the venous thromboembolism event.
‡ Adjusted for the Acute Physiology and Chronic Health Evaluation II score, history of personal or family venous thromboembolism, need for vasopressors, and end-stage renal failure.

TABLE 3. Additional Sensitivity Analyses Using Treatment as a Time-Dependent Covariate in the Cause-Specific Analysis and the Fine and Gray Model

	PLDVT (n = 205)		PE (n = 67)	
Method	**Statistics* (95% CI)**	***P*-Value**	**Statistics*(95% CI)**	***P*-Value**
Univariate analysis				
Fine and Gray model	0.86 (0.53–1.40)	0.54	0.47 (0.25–0.90)	0.02
Cause-specific analysis	0.85 (0.52–1.37)	0.50	0.48 (0.25–0.91)	0.02
Multivariable analysis†				
Fine and Gray model	0.87 (0.53–1.42)	0.58	0.48 (0.26–0.91)	0.02
Cause-specific analysis	0.87 (0.54–1.42)	0.58	0.48 (0.24–0.91)	0.03

CI = confidence interval, PE = pulmonary embolism, PLDVT = proximal leg deep vein thrombosis.
* Sub-hazard ratio (SHR) for the Fine and Gray model; hazard ratio (HR) for the cause-specific analysis.
† Adjusted for the Acute Physiology and Chronic Health Evaluation II score, history of personal or family venous thromboembolism, need for vasopressors, and end-stage renal failure.

PE (Table 4). There were 3659 and 3127 patients included for as-treated and per-protocol analysis, respectively. No significant relationship between dalteparin and decreased risk of PLDVT was found (*P*-values > 0.50): SHR = 0.92 and 0.96 in as-treated and per-protocol analysis for the Fine and Gray model; HR = 0.91 and 0.95 in as-treated and per-protocol analysis for the cause-specific method. Nevertheless, compared with UFH, a superior effect of dalteparin on PE was observed (*P*-values < 0.05), with a SHR of 0.54 and 0.61 in as-treated and per-protocol analysis for the competing risk analysis and a HR of 0.48 and 0.54 in as-treated and per-protocol analysis for the cause-specific method, respectively (Table 4).

DISCUSSION

In this study based on data from the international PROTECT trial, we conducted a competing risk analysis to evaluate the effect of dalteparin versus UFH for the VTE prevention in medical-surgical critically ill patients, taking death as a competing risk. The competing risk analysis showed no significant effect of dalteparin compared with UFH on PLDVT, but a lower risk of PE. These findings were in agreement with results from the cause-specific analysis in the main report.[8] Similar results from additional sensitivity analyses supported the robustness of these findings.

In both the competing risk analysis and cause-specific analysis, we observed a significant difference in PE while no difference in PLDVT for dalteparin compared with UFH. One hypothesis may be due to the difference in nonleg DVT.[8] Nevertheless, there was no significant difference in nonleg DVT between the dalteparin and UFH group.[21] Another possible interpretation may rely on the fact that, unlike the PLDVT, the PE outcome was not screened twice-weekly.[8,12] However, little was known whether the difference in the detection would lead to the difference in the PLDVT and PE outcomes in the PROTECT. More evidence is needed to further explore and clarify the difference in PE for dalteparin versus UFH in critically ill patients.[8,21]

Given the presence of competing risk of death, it may not be appropriate in general to simply censor patients who died before they had a chance to experience a VTE using a cause-specific analysis. Theoretically, the distribution of time-to-censorship may provide information about the distribution of time-to-event, and therefore the assumption of noninformative

TABLE 4. Additional Sensitivity Analyses Using As-Treated (n = 3659) and Per-Protocol (n = 3127) Multivariable Analyses* in PLDVT and PE in the Cause-Specific Analysis and the Fine and Gray Model

	As-Treated Analysis				Per-Protocol Analysis			
Outcome	**Dalteparin (n = 1827)**	**UFH (n = 1832)**	**Statistics* (95% CI)**	***P*-Value**	**Dalteparin (n = 1566)**	**UFH (n = 1561)**	**Statistics* (95% CI)**	***P*-Value**
PLDVT								
Fine and Gray model	94 (5.15)†	108 (5.90)†	0.92 (0.70–1.21)	0.56	91 (5.81)†	99 (6.34)†	0.96 (0.72–1.27)	0.76
Cause-specific analysis			0.91 (0.68–1.23)	0.54			0.95 (0.70–1.29)	0.75
PE								
Fine and Gray model	22 (1.20)†	42 (2.29)†	0.54 (0.33–0.91)	0.02	22 (1.40)†	37 (2.37)†	0.61 (0.38–0.97)	0.04
Cause-specific analysis			0.48 (0.27–0.84)	0.01			0.54 (0.30–0.95)	0.03

CI = confidence interval, PE = pulmonary embolism, PLDVT = proximal leg deep vein thrombosis, UFH = unfractionated heparin.
* Adjusted for the Acute Physiology and Chronic Health Evaluation II score, history of personal or family venous thromboembolism, need for vasopressors, and end-stage renal failure.
* Sub-hazard ratio (SHR) for the Fine and Gray model; hazard ratio (HR) for the cause-specific analysis.
† Expressed as the number and percentage (%) of the venous thromboembolism event.

censoring may not be satisfied in standard survival analysis.[2,16,22] Similarly, those patients who died without having developed a VTE may not be representative of the other patients who remained in the risk set, thereby violating the assumption of independence of survival times between VTE and death.[2,16] Ignoring the competing risk could result in incorrect estimation of the actual risk of VTE in the Kaplan–Meier method,[23–25] and bias the benefit of interventions in trials or the associations between risk factors and outcomes in cohort studies using a cause-specific analysis.[26–28] In contrast, the Fine and Gray model modifies the risk sets such that patients experiencing the competing event are retained artificially in the cohort, with decreasing weight to account for the declining observability in the analyses, rather than being simple censored.[13] The Fine and Gray model could directly use the CIFs to calculate the hazards, and subsequently investigate the treatment effect expressed as a SHR.[16] Compared with the standard survival analysis, the Fine and Gray model had been supported by emerging evidence to account for competing risk, and the SHR had been justified as a better way to estimate the treatment effect than a HR in the Cox regression model.[2,3,14]

In this study, we observed a virtually identical treatment effect of dalteparin versus UFH on VTE in the competing risk analysis and the cause-specific method (Table 2). One of the most important reasons was that the cumulative incidences of mortality between the dalteparin group and UFH group were very similar (Figure 2), in which the 2 cumulative incidence curves overlapped substantially (P-value = 0.23 for log-rank test). Therefore, the similar mortality between the 2 groups yielded analogous censoring in the cause-specific analysis that ignored competing risk. All these findings further supported a similar effect of dalteparin versus UFH on the development of PLDVT but a protective effect on PE in medical-surgical critically ill patients.

There were 3 other trials investigating the efficacy of low-molecular-weight heparin in DVT compared with UFH in medical-surgical critically ill patients,[29–31] as summarized in a recent systematic review.[7] No protective effect of low-molecular-weight heparin on DVT was found in these trials. However, none of them took the competing risk such as death into consideration, despite the high mortality during follow-up. Therefore, it is uncertain whether the competing events would influence the treatment effect reported in these trials.

One study compared the efficacy of dalteparin versus oral anticoagulant therapy in the prevention of recurrent VTE in cancer patients using the standard survival analysis and the Fine and Gray model.[32] These investigators reported a similar treatment effect of dalteparin from these 2 analyses, with a HR of 0.48 in the cause-specific method and a SHR of 0.47 in the competing risk analysis, respectively, and concluded that if the time distribution of competing risks was similar between the treatment groups, standard survival analysis and competing risk method would produce similar findings.[32] If, however, the trial intervention had a different effect on the mortality and the censoring of a competing risk exerted a different influence on the probability of outcomes of interest, a cause-specific analysis ignoring competing risk would lead to misleading findings.[28,32]

For instance, in Pintilie example, he modified the data and assessed the effect of radiation versus chemotherapy on cardiac hospitalization in 689 patients with Hodgkin lymphoma.[3] The cause-specific analysis yielded an effect of radiation versus chemotherapy on cardiac hospitalization that was not significant (HR: 1.07, 95% CI: 0.71–1.63), while the Fine and Gray model found that radiation was significantly related with increased risk of cardiac hospitalization (SHR: 1.63, 95% CI: 1.10–2.45). The interpretation relied on the fact that the treatment groups had different effect on the risk of death (HR: 0.38, 95% CI: 0.31–0.47 for radiation versus chemotherapy). In other words, the time distribution of death in the chemotherapy group was different from the radiation group, and more patients died in the chemotherapy group before they could experience cardiac hospitalization than in the radiation group. Even though no clinical conclusion could be drawn due to the modification of the data,[3] this example did show that the cause-specific method ignoring the difference of censoring of the competing risk would result in a biased finding. Therefore, in the presence of competing events, a competing risk analysis or a comparison between a competing risk and standard survival analysis would be recommended to minimize the potential impact of competing risks and avoid incorrect conclusions.

Strengths of this study include a large sample size from a multicenter RCT and the use of optimally available statistical methods to investigate the efficacy of dalterparin versus UFH in VTE, taking into death as a competing risk account. Additional sensitivity analyses were also conducted to further assess and support the robustness of the original findings. However, we did not account for the transfer to a nontrial hospital as another potential competing event for VTE in this study. Given the limited data recorded in the database, no analysis could be performed to assess whether the competing risk of transfer to a non-trial hospital would impact the estimation of treatment effect on VTE. Furthermore, because this study focused on methodological analysis and aimed to assess the impact of competing risk, we did not have the data on serological tests and coagulation states for patients in this RCT. Therefore, for the phenomenon that there was significant difference in PE but no difference in PLDVT in the treatment groups, we could not use the serological or coagulation results to further illuminate the mechanism.

CONCLUSION

In this competing risk analysis using data from an international critical care trial, no significant difference was found between dalteparin and UFH on PLDVT, but dalteparin significantly reduced the risk of PE in medical-surgical critically ill patients. All the findings from the competing risk method were in accordance with results from a cause-specific analysis, increasing the inferences from the original findings.

ACKNOWLEDGMENTS

The authors thank the contributions of the patients, research coordinators, and bedside staff in the PROTECT herein.

G. Li: contributed to the study conception and design, data analysis, interpretation of the data and drafting the manuscript. D. Cook, and M. Levine: contributed to the study design, interpretation of the data, critical revision of the manuscript, and approval of the final version of the manuscript. G. Guyatt, and M. Crowther: contributed to interpretation of the data, critical revision of the manuscript, and approval of the final version of the manuscript. D. Heels-Ansdell: contributed to the analysis and interpretation of the data, critical revision of the manuscript, and approval of the final version of the manuscript. A. Holbrook, F. Lamontagne, and S. Walter: contributed to interpretation of the data, critical revision of the manuscript, and approval of the final version of the manuscript. N. Ferguson, S. Finfer, Y. Arabi, R. Bellomo and D. Cooper: contributed to critical version of the manuscript and approval of the final version of the manuscript. L. Thabane: contributed to the study

design, interpretation of the data, critical revision of the manuscript, and approval of the final version of the manuscript.

The sponsors had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

PROTECT Collaborators: Canada Investigators—Sherbrooke University Hospital and Centre de Recherche Clinique Étienne-Le Bel, Sherbrooke: Olivier Lesur, Francois Lamontagne, Sandra Proulx, Sylvie Cloutier, Brigitte Bolduc, Marie-Pierre Rousseau, Julie Leblond, Gérard Schmutz; Ottawa General Hospital, Ottawa: Lauralyn McIntyre, Paul Hebert, IreneWatpool, TracyMcArdle, Claude Gaudert, Paule Marchand, Carson Davidson, Anne-Marie Dugal, Susan Fetzer, Wael Shabana, Marc Castonguay, Sohail Anwar, Valentina Kozarenko, Shahina Mohammad, Svitlana Sikalska, Suzanne Gauthier, Arif Mustafa; St Paul's Hospital, Vancouver: Peter Dodek, Betty Jean Ashley, Sheilagh Mans, Mara Pavan, Jonathon Leipsic, Sam Meiersdorf, Adrian Yoong, Francisco Avila Flores, Hina Mumtaz, Patrick Ng, Cathy Fix; Hamilton Health Sciences, Hamilton General Hospital, Hamilton: Maureen Meade, Lori Hand, Maya Biljan, Michael Patlas, Lianne Broughton, Lucy Degrow, Dianna Connor,Maggie Tuhy, Dawn Whyte, Meaghan Jefferson, Kaitlyn Aarts, Lindsay Vooys, Michael Anzovino; Radiology Department, Maisonneuve Rosemont Hospital, Montreal: Yoanna Skrobik, Johanne Harvey, Stefania Chitu, Marceline Quach, Linda Pinet; Kingston General Hospital, Kingston: John Muscedere, Susan Fleury, Nicole Godfrey, Sharlene Hammond, Elizabeth Mann, Monica Myers, Amber Robinson, Chris Grey, Eric Saurbrei, Jennifer Cox, Angela Nugent, Julie Kolesar, Amy Fisher, Amy Northrup-St-Onge, Marshaw Paterson-Skeete, Wendy Schlottke, Wendy Bertrim, Cathy Marshall; St Joseph's Healthcare, Hamilton: Deborah Cook, Ellen McDonald, Andrea Tkaczyk, France Clarke, Christine Wallace, David Schiff, Jennifer McDonald, Sarah Todd, Patty Harkness, Angela Medic, Joanna Andrews, Moira Sands, Iwona Hall, Tanya Boniakowski, Kim Lichty. Australia Investigators—Nepean Hospital, Sydney: Ian Seppelt, LeonieWeisbrodt, Robyn Bond, Stella Suen, Jason Trinh, Roger Hall, Richard Huang, Helen Chow; Blacktown Hospital, Blacktown: Graham Reece, Treena Sara, Kiran Nand, Rabsima Ibrahim, James Jarrett, Jagdish Seehra, Gill Stringer. Saudi Arabia Investigators—King Abdulaziz University Hospital, Jeddah: Jamal Alhashemi, Sanaa Shalabi, Randa Ainosah, Julie Ann Sonbul, Rustico Gloriani, Rosalinda Huertazuela, Ibrahim Abbas, Judy Chavez, Nahid El Toum; King Saud Bin Abdulaziz University for Health Sciences, Riyadh: Mohammed Alsultan, Yaseen Arabi, Riette Brits. Brazilian Investigators—Santa Casa Hospital, Porto Allegre: Marcelo Garcia da Rocha (Co-Lead), Andréa Kramer, Martha Hädrich; Hospital Coracao Research Institute HCor, São Paulo: Otavio Berwanger, Edson Romano, Anna Maria Buehler. United Kingdom Investigators—King's College London, Guy's and St Thomas' Hospital, London: Marlies Ostermann, David Treacher, Tony Sherry, John Smith, Barnaby Sanderson, Josephine Ng, John Brooks, Ling Lim, Katie Lei, Paul Tunstell, Cathy McKenzie, Francesco Cicirello, T. S. Padayachee, Nicholas Thomas, Andrew. J. Arnold. United States Investigators—Mayo Clinic, Rochester: Nicholas E. Vlahakis, Laurie Meade, Debbie Bauer, Bradley Lewis, Nora Harer.

REFERENCES

1. Gooley TA, Leisenring W, Crowley J, et al. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med.* 1999;18:695–706.
2. Putter H, Fiocco M, Geskus R. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med.* 2007;26:2389–2430.
3. Pintilie M. An introduction to competing risks analysis. *Revista Española de Cardiología (English Edition).* 2011;64:599–605.
4. Geerts WH, Bergqvist D, Pineo GF, et al. Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133:381S–453S.
5. Cook D, Crowther M, Meade M, et al. Deep venous thrombosis in medical-surgical critically ill patients: prevalence, incidence, and risk factors. *Crit Care Med.* 2005;33:1565–1571.
6. Geerts W, Selby R. Prevention of venous thromboembolism in the ICU. *Chest J.* 2003;124:357S–363S.
7. Alhazzani W, Lim W, Jaeschke RZ, et al. Heparin thromboprophylaxis in medical-surgical critically ill patients: A Systematic Review and Meta-Analysis of Randomized Trials*. *Crit Care Med.* 2013;41:2088–2098.
8. Cook D, Meade M, Guyatt G, et al. Dalteparin versus unfractionated heparin in critically ill patients. *N Engl J Med.* 2011;364:1305–1314.
9. Forsblom C, Harjutsalo V, Thorn LM, et al. Competing-risk analysis of ESRD and death among patients with type 1 diabetes and macroalbuminuria. *J Am Soc Nephrol.* 2011;22:537–544.
10. Berry SD, Ngo L, Samelson EJ, et al. Competing risk of death: an important consideration in studies of older adults. *J Am Geriatr Soc.* 2010;58:783–787.
11. Resche-Rigon M, Azoulay E, Chevret S. Evaluating mortality in intensive care units: contribution of competing risks analyses. *Crit Care.* 2005;10:R5.
12. Cook D, Meade M, Guyatt G, et al. PROphylaxis for ThromboEmbolism in Critical Care Trial protocol and analysis plan. *J Crit Care.* 2011;26:223e221-223.e229.
13. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999;94:496–509.
14. Bakoyannis G, Touloumi G. Practical methods for competing risks data: a review. *Stat Methods Med Res.* 2012;21:257–272.
15. Thabane L, Mbuagbaw L, Zhang S, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol.* 2013;13:92.
16. Kleinbaum DG, Klein M. Survival Analysis – A Self-Learing Text. 3rd ed. Springer Science and Business Media, LLC, New York, NY, USA: Springer; 2012.
17. Pepe MS, Mori M. Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med.* 1993;12:737–751.
18. Cook D, Crowther M, Meade M, et al. Deep venous thrombosis in medical-surgical critically ill patients: prevalence, incidence, and risk factors. *Crit Care Med.* 2005;33:1565–1571.
19. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994;81:515–526.
20. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Ann Rev Public Health.* 1999;20:145–157.
21. Lamontagne F, McIntyre L, Dodek P, et al. Nonleg venous thrombosis in critically ill adults: a nested prospective cohort study. *JAMA Inter Med.* 2014;174:689–696.
22. Satagopan J, Ben-Porat L, Berwick M, et al. A note on competing risks in survival data analysis. *Br J Cancer.* 2004;91:1229–1235.
23. Ay C, Posch F, Kaider A, et al. Estimating risk of venous thromboembolism in patients with cancer in the presence of competing mortality. *J Thromb Haemost.* 2015;13:390–397.

24. Melberg T, Nygård OK, Kuiper KK-J, et al. Competing risk analysis of events 10 years after revascularization. *Scand Cardiovasc J.* 2010;44:279–288.

25. Hsieh KP, Chen LC, Cheung KL, et al. A competing risk analysis of hormone therapy interruption in Asian women with breast cancer. *Pharmacoepidemiol Drug Saf.* 2015;24:301–309.

26. Koller MT, Raatz H, Steyerberg EW, et al. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med.* 2012;31:1089–1097.

27. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol.* 2012;41:861–870.

28. Berry SD, Ngo L, Samelson EJ, et al. Competing risk of death: an important consideration in studies of older adults. *J Am Geriatr Soc.* 2010;58:783–787.

29. Shorr AF, Williams MD. Venous thromboembolism in critically ill patients. Observations from a randomized trial in sepsis. *Thromb haemost.* 2009;101:139–144.

30. De A, Roy P, Garg VK, et al. Low-molecular-weight heparin and unfractionated heparin in prophylaxis against deep vein thrombosis in critically ill patients undergoing major surgery. *Blood Coagul Fibrinolysis.* 2010;21:57–61.

31. Goldhaber SZ, Kett DH, Cusumano CJ, CAK. Low molecular weight heparin versus minidose unfractionated heparin for prophylaxis against venous thromboembolism in medical intensive care unit patients: a randomized controlled trial (Abstract). *J Am Coll Cardiol.* 2000;35(Suppl):325A.

32. Parpia S, Julian J, Thabane L, et al. Competing events in patients with malignant disease who are at risk for recurrent venous thromboembolism. *Contemp Clin Trials.* 2011;32:829–833.

e-Table 1. Details of PROTECT Research Ethics Board (REB) approvals

Institution	Name of REB	REB Number
St Joseph's Hospital	St Joseph's Healthcare Hamilton REB	05-2572
QEII, Halifax	Capital Health Research Ethics Board	2005-321
Hamilton Health Science-Hamilton General Hospital	Hamilton Health Science Research Ethics Board	05-429
Hamilton Health Science-McMaster Medical Center	Hamilton Health Science Research Ethics Board	05-429
University Health Network-Toronto General Hospital	University Health Network Research Ethics Board	05-0934-A
Centre hospitalier affi llie universitaire de Quebec-Enfant Jesus Hospital	Centre hospitalier affi llie universitaire de Quebec-Research Ethics Board	PEJ-338
Hopital Charles LeMoyne	Comite d'ethique de la recherché Hopital Charles LeMoyne	2005-06-52
Ottawa Hospital-Civic Campus	The Ottawa Hospital Research Ethics Board	2005872-01H
Ottawa Hospital-General Campus	The Ottawa Hospital Research Ethics Board	2005872-01H
Hopital Maisonneuve Rosemont	Comite d'ethique de la recherché de l'hopital Maisonneuve Rosemont	#05093
Hopital du Sacre Coeur de Montreal	Comite d'ethique de la recherché Sacre Couer Hospital	C.E 2005-12-80
Mount Sinai Hospital	Mount Sinai Hospital Research Ethics Board	05-0259-A
Sunnybrook Health Science Center	Sunnybrook Health Science Center Research Ethics Board	043-2006
St Michael's Hospital	S t Michael's Hospital Research Ethics Board	06-007
Kingston General Hospital	Queens University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board	DMED-915-05
Hospital Moinhos de Vento	Comite de Etica em Pesquisa e Comissao Cientifi ca Hospital Moinhos de Vento	CEP-IEP HMV:2006/37
Hospital ProCardiaco	Comite de Etica em Pesquisa e Comissao Cientifi ca Hospital ProCardiaco	CEP 236
Royal Columbian Hospital	Fraser Health Research Ethics Board	FHREB 2005-96
Vancouver General	The University of British Columbia Clinical	H05-70623

Hospital	Research Ethics Board	
St Paul's Hospital	The University of British Columbia Providence Health Care Research Ethics Board	H05-50274
Foothills Hospital	Conjoint Health Ethics Research Board	E-20147
Royal Alexandra Hospital	The Health Research Ethics Board	Pro00001957
University Health Network-Toronto Western Hospital	University Health Network Research Ethics Board	05-0934-A
The Alfred Hospital	The Alfred Ethics Committee	236/05
Royal Melbourne Hospital	Research Directorate-Human Ethics Committee, Melbourne Health	2005.224
MD Anderson Cancer Center	MD Anderson Cancer Center Institutional Review Board	2008-0466
Irmandade da Santa Casa de Misericordia de Porto Alegre	Comite de Etica em Pesquisa e Comissao Cientifi ca Irmandade da Santa Casa de Misericordia de Porto Alegre	1368/06
Hospital do Coracao	Comite de Etica em Pesquisa e Comissao Cientifi ca Hospital do Coracao	036/2007
St John's Mercy Medical Center	St John's Mercy Medical Center Institutional Review Board	09-021
L'Hopital Laval	Comite d'ethique de la recherché de l'hopital Laval	20183
Hamilton Health Science-Henderson Hospital	Hamilton Health Science Research Ethics Board	05-429
Hospital Sao Paulo	Comite de Etica em Pesquisa Hospital Sao Paulo	CEP 1523/06
University of Alberta Medical Research Center	Health Research Ethics Board	6174
Mayo Clinic	Mayo Clinic Institutional Review Board	08-002579
Austin Health	Austin Health Human Research Ethics Committee	H2006/02436
Boxhill Hospital	Eastern Health Research and Ethics Committee	E70/0506
Frankston Hospital	Peninsula Health Research and Ethics Committee	2005-59
Monash Medical Center	Southern Health Research Ethics Committee	05173B
Nepean Hospital	Human Research Ethics Committee, Nepean Campus	06/030
Royal Adelaide Hospital	Royal Adelaide Hospital Research Ethics Committee	070323

King Faisal and Specialist Research Center	King Faisal and Specialist Research Center –Institutional Review Board	2006-02
Royal Prince Alfred Hospital	Sydney South West Area Health Service Ethics Review Committee	X06-0047
Peter Lougheed Hospital	Conjoint Health Ethics Research Board	E-20147
Bendigo Health Center	Bendigo Health Human Research Ethics Committee	1/2006
Blacktown Hospital	Human Research Ethics Committee, Nepean Campus	06/030
Flinders Medical Center	Flinders Clinical Research Ethics Committee	120/056
Geelong Hospital	Barwon Health Research and Ethics Advisory Committee	06/04
McGill University Health Center-Montreal General	Center Universitaire de Sante McGill Bureau d'ethique de la Recherche	BMB 07-00
Riyadh Military Hospital	Riyadh Military Hospital Research and Ethical Committee	338-2008
Dandenong Hospital	Southern Health - Human Research Ethics Committee B	05173B
King Fahad Medical City	Institutional review Board King Fahad Medical City	004-08
University of Wollongong	University of Wollongong South Eastern Sydney Illawarra NSW Health Human Research Ethics Committee	236/05
Vancouver Island Health Authority	Research Review and Ethical Approval Committee	2006-21
Guy's and St Thomas Hospital	Guy's and St Thomas Hospital National Foundation Trust Research and Development	09/H0802/022
Center Hospitalier Universitaire de Sherbrook	Comite d'ethique de la recherché en Santechez l'Human	07-019
McGill University Health Center-Royal Victoria Hospital	Center Universitaire de Sante McGill Bureau d'ethique de la Recherche	BMB 07-003
Guelph General Hospital	Guelph General Hospital Research Ethics Board	7-010
Grand River Hospital	Tri-Hospital Research Ethics Board	07-150
St Boniface Hospital	Bannatyne Campus Research Ethics Board	B2007:085
Lakeridge Health	Lakeridge Health Research Ethics Board	2008-007

Rhode Island Hospital	Lifespan Institutional Review Board	00004624
Lyell McEwin Hospital	Southern Health Research Ethics Committee	05173B
Surrey Memorial Hospital	Fraser Health Research Ethics Board	FHREB 2005-96
King Abdulaziz Medical City Hospital	King Abdulaziz Medical City Hospital Institutional Review Board	RC07/015
London Health Science Center-University Campus	University of Western Ontario Research Ethics Board	16565
King Abdulaziz University Hospital	King Abdulaziz University Hospital Bioethical and Research Committee	225-08

Chapter 6

CONCLUSIONS

In this thesis, we have focused on three methodological issues in prediction models and data analyses using observational and clinical trial data: 1). prediction of individual combined benefit and harm outcomes related to warfarin therapy, 2). risk factors for and prediction of mortality in critically ill medical-surgical patients receiving heparin thromboprophylaxis, and 3). competing risk analysis for evaluation of dalteparin versus unfractionated heparin for venous thromboembolism in medical-surgical critically ill patients. These three important topics are investigated in this manuscript-basis thesis, with each study dedicated to exploring each of the issues. This chapter summarizes the key findings from Chapters 2 to 5, and discusses the implications and limitations of the three studies.

In **Chapter 2** and **3**, we used a new approach to build a prediction model of individual combined benefit and harm outcomes related to warfarin therapy in patients diagnosed with AF, based on data from an anticoagulation management cohort. We utilized a polytomous logistic regression (PLR) model to identify risk factors and then construct the new prediction model. In the study design and model building, we controlled the potential of immortal time bias and assessed the impact of competing risk of death on combined benefit and harm outcomes. Model performances and validation were evaluated systematically in the study. The prediction model was a good fit, had acceptable discrimination and calibration, and was internally and externally validated.

<u>Implications of this study</u>

There are a large number of clinical prediction models for risk of stroke and major bleeding in patients with AF. Among the more prominent and commonly used prediction models are $CHADS_2$/CHA_2DS_2-VASc for risk of stroke, and HAS-BLED for risk of major bleeding. Nevertheless, these risk-stratification scores cannot estimate individual combined benefit and

harm probabilities simultaneously when the initiation of warfarin therapy is considered. A prediction model that provides comprehensive information on the individual patient-specific probabilities of both benefit and (decreased risk of stroke) and harm (increased risk of bleeding) is needed to aid in patient-physician shared decision-making when they are considering warfarin therapy. In addition, some studies tried to use the 'net benefit' method[1-3] or the 'combined stroke and bleeding risk scores' approach[4, 5] to take into account benefit and harm outcomes at the same time. However, concerns about the 'net benefit' approach included the weighting factor and the exclusion of GI bleeding in its calculation. Moreover, the 'combined stroke and bleeding risk scores' approach could not improve prediction of stroke and major bleeding beyond the individual stroke ($CHADS_2$, CHA_2DS_2-VASc) or bleeding scores (HAS-BLED).[6] By contrast, our new prediction model may be more practical and acceptable in real-world settings because the PLR model can provide specific probabilities of stroke and major bleeding at the same time at the individual-level. The new prediction model may aid in patient-physician shared decision-making regarding warfarin therapy initiation by offering more personalized and detailed estimates of combined benefit and harm outcome probabilities. Should this approach be validated in other patient populations, it has potential advantages over existing risk stratification approaches. More research is needed to further explore the prediction model's external validity and feasibility in clinical practice.

Limitations of this study

Though we used a large sample of patients, tried to control potential biases and conducted rigorous statistical analyses, several limitations exist in this study. The data accuracy of baseline and outcome information may be less than optimal, because of no confirmatory chart review to validate the data. This would weaken and impair the findings based on the data. Another limitation is that we could only focus on three outcomes (stroke with no major bleeding, major bleeding with no stroke, and neither event) due to insufficient sample size, though we intended to predict the fourth outcome, i.e., both stroke and major bleeding, as well. Furthermore, since there are no other contemporary cohorts for external validation, the generalizability of the new prediction model is uncertain.[7] Therefore more studies are needed

to validate the prediction model in other populations and settings. Besides, we did not consider the classification and regression tree (CART) method that was another commonly-used approach for model building, given the simplicity and greater discrimination and predictive accuracy of the PLR method shown in other studies.[8-11] It would be another interesting topic to compare model performances between the CART and PLR approaches in predicting combined benefit and harm outcomes. Likewise, it may be also worth attempting other approaches such as random forest regression and support vector machines in prediction of combined benefit and harm outcomes related to warfarin in AF.

In **Chapter 4**, we identified risk factors independent of APACHE II score and constructed a new prediction model for risk of mortality in critically ill medical-surgical patients receiving thromboprophylaxis, using data from an international thromboprophylaxis trial. The new model was a good fit, well calibrated and internally validated. Nevertheless, the discriminative ability of the prediction model was not satisfactory.

Implications of this study

The predictive accuracy of the APACHE II system for risk of mortality in critically ill patients has been found to be limited in different populations and countries.[12] In this study using data from a multicenter thromboprophylaxis trial, the APACHE II score had surprisingly low discrimination in prediction of risk of hospital mortality in critically ill medical-surgical patients. Though our new prediction model had significantly higher discrimination than the APACHE II score alone, adding more information increased the discriminative ability to only a small extent. Compared with the APACHE II score alone, the new prediction model increased data collection, was more complex but did not substantially improve discriminative ability. Therefore the utilization of the new prediction model would be limited to situations where a clinician or health services investigator was sufficiently dissatisfied with the APACHE II score and was requiring an even minimally better model to predict risk of death in critically ill patients. More studies are warranted to further explore how to improve discriminative ability of the predictions in risk of mortality in critically ill patients.

Limitations of this study

The generalizability of the findings in this study would be limited because the data used were from a randomized trial with strict inclusion and exclusion criteria for participants. Moreover, only the data documented in the original trial database could be used in building the prediction model. No other potentially important indicators of illness severity could be captured, which therefore may lead to the low discrimination of the prediction model in this study. Regarding the model building, we first pruned the candidate predictors based on the criterion of variance inflation factor (VIF). However, except for VIF, other factors including low sample size and large unexplained variance may also lead to model unstability and poor performance.[13] Furthermore, we chose the backward elimination for variable selection to construct the model accordingly to the recommendation in the literature.[14, 15] It still remained unclear whether other variable selection methods would outperform backward elimination in this study.

In **Chapter 5**, we performed a competing risk analysis as a sensitivity analysis to assess the efficacy of dalteparin versus unfractionated heparin in preventing VTE in medical-surgical critically ill patients, taking all-cause death as a competing risk for VTE. The competing risk analysis yielded similar results from the cause-specific analysis in the main report;[16] that is, no significant effect of dalteparin compared with unfractionated heparin on PLDVT, but a lower risk of pulmonary embolism was found.

Implications of this study

Due to the ever-present competing risk of death, it may not be appropriate to simply censor patients who died before they had a chance to develop a VTE using a cause-specific analysis because the assumptions of non-informative censoring and independence of time distributions between VTE and death may be violated. The distribution of time-to-censorship may provide information about the distribution of time-to-event, and therefore the assumption of non-informative censoring may not be satisfied in cause-specific analysis.[17-19] Likewise, the patients who died without having experienced a VTE may not be representative of the other

patients who remained in the risk set, thereby infringing the assumption of independence of survival times between VTE and death.[17, 18] Thus ignoring the competing risk would bias the treatment effect estimates in trials in a cause-specific analysis. In this study, a virtually identical treatment effect of dalteparin versus unfractionated heparin on VTE in the competing risk analysis and the cause-specific method was found. One important reason was that the cumulative incidences of mortality between the dalteparin group and unfractionated heparin group were very similar. Hence, the similar mortality between the two groups yielded analogous censoring in the cause-specific analysis that ignored competing risk. Nevertheless, if the interventions had a different effect on the mortality and the censoring of a competing risk had a different influence on the probability of outcomes of interest, a cause-specific analysis ignoring competing risk would lead to biased findings.[20, 21] Therefore, in the presence of competing events, a competing risk analysis or a comparison between a competing risk and cause-specific analysis would be recommended to minimize the potential impact of competing risks and avoid misleading conclusions.

Limitations of this study

Some limitations exist in this study. We could not take into account the transfer to a non-trial hospital as another potential competing event for VTE, because of the limited data recorded in the database. Moreover, no data on serological tests and coagulation states for patients in this study were available. Thus for the phenomenon that there was a significant difference in pulmonary embolism but no difference in PLDVT between dalteparin group and unfractionated heparin group, we could not use the serological or coagulation results to further illuminate the mechanism.

In summary, this sandwich thesis investigated methodological issues in prediction models and data analyses using observational and clinical trial data. The three projects contribute to the literature by providing a new approach to considering combined benefit and harm outcomes related to warfarin therapy in prediction models, by developing a new prediction model after combining APACHE II score and other risk factors in prediction of risk of mortality in

critically ill patients, and by assessing the impact of competing risk on treatment effect estimates in clinical trials. More investigation may be needed to generalize these findings and explore further methodological issues in prediction models and data analyses.

References:

1. Singer DE, Chang Y, Fang MC, Borowsky LH, Pomernacki NK, Udaltsova N, Go AS. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. *Ann Intern Med*. 2009;151:297-305
2. Friberg L, Rosenqvist M, Lip GY. Net clinical benefit of warfarin in patients with atrial fibrillation: A report from the swedish atrial fibrillation cohort study. *Circulation*. 2012;125:2298-2307
3. Olesen JB, Lip GY, Lindhardsen J, Lane DA, Ahlehoff O, Hansen ML, Raunso J, Tolstrup JS, Hansen PR, Gislason GH, Torp-Pedersen C. Risks of thromboembolism and bleeding with thromboprophylaxis in patients with atrial fibrillation: A net clinical benefit analysis using a 'real world' nationwide cohort study. *Thromb Haemost*. 2011;106:739-749
4. Lip GY, Lane DA, Buller H, Apostolakis S. Development of a novel composite stroke and bleeding risk score in patients with atrial fibrillation: The amadeus study. *Chest*. 2013;144:1839-1847
5. Banerjee A, Fauchier L, Bernard-Brunet A, Clementy N, Lip GY. Composite risk scores and composite endpoints in the risk prediction of outcomes in anticoagulated patients with atrial fibrillation. The loire valley atrial fibrillation project. *Thromb Haemost*. 2014;111:549-556
6. Dzeshka MS, Lane DA, Lip GY. Stroke and bleeding risk in atrial fibrillation: Navigating the alphabet soup of risk-score acronyms (chads2 , cha2 ds2 -vasc, r2 chads2 , has-bled, atria, and more). *Clinical cardiology*. 2014;37:634-644
7. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: Validating a prognostic model. *BMJ: British Medical Journal*. 2009;338:1432-1435
8. Bilen K, Ponzer S, Ottosson C, Castren M, Pettersson H. Deliberate self-harm patients in the emergency department: Who will repeat and who will not? Validation and development of clinical decision rules. *Emergency medicine journal : EMJ*. 2013;30:650-656
9. Ferreira D, Oliveira A, Freitas A. Applying data mining techniques to improve diagnosis in neonatal jaundice. *BMC medical informatics and decision making*. 2012;12:143
10. Ho WH, Lee KT, Chen HY, Ho TW, Chiu HC. Disease-free survival after hepatic resection in hepatocellular carcinoma patients: A prediction approach using artificial neural network. *PloS one*. 2012;7:e29179
11. Minne L, Eslami S, de Keizer N, de Jonge E, de Rooij SE, Abu-Hanna A. Statistical process control for monitoring standardized mortality ratios of a classification tree model. *Methods of information in medicine*. 2012;51:353-358
12. Siontis G, Tzoulaki I, Ioannidis J. Predicting death: An empirical evaluation of predictive tools for mortality. *Archives of internal medicine*. 2011:archinternmed. 2011.2334 v2011
13. O'brien RM. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*. 2007;41:673-690
14. Mantel N. Why stepdown procedures in variable selection. *Technometrics*. 1970;12:621-625
15. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ (Clinical research ed.)*. 2009;338
16. Cook D, Meade M, Guyatt G, Walter S, Heels-Ansdell D, Warkentin TE, Zytaruk N, Crowther M, Geerts W, Cooper DJ. Dalteparin versus unfractionated heparin in critically ill patients. *The New England journal of medicine*. 2011;364:1305-1314
17. Kleinbaum DG, Klein M. *Survival analysis- a self-learing text*. Springer 2012.
18. Putter H, Fiocco M, Geskus R. Tutorial in biostatistics: Competing risks and multi‐state models. *Statistics in medicine*. 2007;26:2389-2430
19. Satagopan J, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach A. A note on competing risks in

survival data analysis. *British journal of cancer*. 2004;91:1229-1235

20. Berry SD, Ngo L, Samelson EJ, Kiel DP. Competing risk of death: An important consideration in studies of older adults. *Journal of the American Geriatrics Society*. 2010;58:783-787

21. Parpia S, Julian J, Thabane L, Lee A, Rickles F, Levine M. Competing events in patients with malignant disease who are at risk for recurrent venous thromboembolism. *Contemporary clinical trials*. 2011;32:829-833