January 2018

Precision Medicine In Atrial Fibrillation: A Risk Model For Management With Dabigatran V. Warfarin

Jeremy Ader

Precision Medicine in Atrial Fibrillation: A Risk Model for Management with Dabigatran v. Warfarin

A Thesis Submitted to the
Yale University School of Medicine
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

by

Jeremy Ader

MD/MBA Class of 2018

Abstract

The Precision Medicine Initiative aims to advance Medicine from "one-size-fits-all" treatments to more individualized approaches. Clinical trials evaluate treatments by analyzing average outcomes, and thus risk overlooking potential differences in treatment effect among different subsets of the study population. The use of multivariate models has been proposed as a way to identify heterogeneity of treatment effect and to determine patients' individualized treatment risks and benefits.

We analyzed the Randomized Evaluation of Long-Term Anticoagulation Therapy (RELY) trial of dabigatran versus warfarin in patients with atrial fibrillation, to determine if the application of multivariate predictive models could demonstrate heterogeneity of treatment effect among the study population. We developed two models to predict patients' risk of stroke or systemic embolism and risk of major bleeding if treated with dabigatran or warfarin. We then applied these models to the individual patients in the RE-LY trial, and determined patients difference in risk if treated with dabigatran versus warfarin. Individual difference in stroke risk for dabigatran 110mg and 150mg versus warfarin was -0.78% ± 0.95% and -1.32% ± 1.31% and the difference in major bleeding risk was -1.12% ± 1.44% and -0.41% ± 2.39%, respectively.

These findings demonstrate heterogeneity of treatment effect in the RE-LY trial, and the ability of multivariate risk models to identify distinct treatment risks for individual patients. Such models could be used in clinical practice provide patients and clinicians with individualized treatment risk information and improve treatment decisions.

Acknowledgements

First, I would like to thank Nihar Desai, MD, MPH, who has been a generous and incredible research, career and life mentor and role model since my first year of medical school. This thesis, and the wonderful opportunity to work on it, would not have been possible without his guidance.

I want to acknowledge, Nihar Desai, John Spertus and Sophie Tang for all of their work on this project. I would also like to thank Harlan Krumholz, and the Center for Outcomes Research and Evaluation for supporting this research. And I would also like to thank the RE-LY investigators, and all of the patients who participated in the RE-LY trial.

To me, the Yale Thesis is a very special culmination of a lot of hard work, a lot of support, and a lot of good fortune. I want to thank the Yale School of Medicine for an incredible medical education, and for providing the opportunity to pursue this thesis project. I would also like to thank all of my teachers from Concord Road Elementary School, Ardsley Middle School, Ardsley High School, Brown University and the Yale School of Medicine and Yale School of Management for giving me the skills and curiosity that went into writing this thesis.

I want to thank my younger brother Ethan and my big sister Melissa for rooting me on all of these years. And I want to thank my wife Nikki, the love of my life, who has guided me and supported me, through five years up and down the New Haven Metro North line.

I am dedicating this thesis to my mother, Marcie Schneider and my father, Michael Ader who held me up and celebrated me through everything. I am where I am, and who I am, because of your love.

Table of Contents

Introduction

In 1970, the famous Boston Red Sox hitter Ted Williams, known best for his extraordinary and yet untouched record .408 batting average, published "the science of hitting," a book to teach boys and girls across America how to consistently hit a baseball.[a] He demonstrated the first rule of hitting, "get a good ball to hit," through an illustration of the strike zone that showed his personal batting average for balls thrown in different locations (1).

[a] Note the author is a fan of the 27-Time World Champion New York Yankees, not the Boston Red Sox

The purpose of this diagram was to show that every individual should calculate their own unique batting average for pitches in different locations, so they could swing at those pitches where they had a high average, and avoid pitches where they didn't hit as well. The notion that one size doesn't fit all, and that unique statistical profiles can inform individualized hitting styles, was one of the earliest and most provocative displays of the power of infusing the art of hitting, with science.

In his 2015 State of the Union Address, President Barack Obama launched the Precision Medicine Initiative, with the mission of advancing Medicine from "on-size-fits-all" treatments to more individualized approaches. Underlying this initiative, is the premise that while treatments are generally developed for the "average patient," most patients are not in fact, average (2). The urgency of this initiative is brought on by a recognition that Medicine is in an age of increasing availability of data from sources such as genomic profiles, advanced diagnostic imaging, wearable devices, and clinical records in electronic health records. This data presents the opportunity to create unique patient profiles that can guide personalized treatment decisions, and to achieve an age of "precision medicine."(3) The use of such profiles has the potential to enhance both the science and the art of Medicine. In this Thesis, I evaluate the potential to use multivariate predictive risk models, to create individualized patient treatment risk and benefit profiles, that can be used by clinicians and patients to make more personalized treatment decisions.

The application of traditional randomized trials to clinical practice and the limitation of subgroup analyses

Randomized controlled trials are the gold standard for evaluating the effectiveness of medical treatments. These trials are designed such that their results are generalizable, and can inform treatment decision for broad populations of patients. Results are usually reported as an overall treatment effect, defined as the average effect observed among all enrolled patients. However, a major limitation of this practice, is that the average effect may be heavily influenced by a subset of the population that has large treatment effect, even if much of the population saw no treatment effect (4). In practice, clinicians need to figure out how to apply these average results to individual patients, each with their own distinct comorbidities and demographics, and consequently, potentially different risks. Clinical investigators commonly address this challenge by publishing subgroup analyses, through which patients are categorized according to an individual variable, and results are published for each patient subgroup. For example, a given trial might publish the results of an antihypertensive for men and for women, for people over the age of 65 and people under the age of 65, or people with diabetes and people without diabetes.

While these subgroup analyses can be helpful, they face a number of challenges. First, they have statistical limitations. They often have very limited statistical power and consequently are subject to false negatives. Additionally, as the number of subgroups increases, false positives will increase and reliability will subsequently decrease. Burke et al explains the issue of false positive findings in subgroup analyses, through the analogy

of ordering diagnostic testing for patients with and without an indication. Following Bayes' rule, positive findings in the group without an indication for testing, are more likely to be false positives, than if they were found in the group who had an indication. Similarly, adding numerous subgroup analyses to a research study, without strong hypotheses for each analysis, is likely to lead to false positive findings (5). Last, subgroup analyses, are limited to analyzing populations by categorical variables, and are unable to take full advantage of continuous data such as age or blood pressure (4).

Second, subgroup analyses can pose dilemmas for individual treatment decision making, especially given their usual reporting of hazard ratios, or relative risk reduction without corresponding information about pre-treatment risk, or absolute risk reduction. For example, suppose a clinical trial compared anticoagulants A and B for stroke prevention in atrial fibrillation, and found drug A to be associated with lower rates of stroke for people over the age of 65, and drug B to be associated with lower rates of stroke for people with diabetes. In such a scenario, how do a 70-year-old diabetic patient and her physician determine which medication would be best? To answer this question, the physician and the patient need to know whether the absolute benefit of stroke prevention of drug A is larger than the absolute benefit of drug B. And to determine this absolute benefit, we need to know the relative risk reduction with drug A and drug B, and the pre-treatment risk of their populations. For example, if the 10-year pre-treatment risk of stroke in the 65 years and older population is 10%, and drug A showed a 50% reduction in stroke, then the absolute decrease in stroke risk from using drug A is 5%. On the other hand, if the 10-year pretreatment risk of stroke in the diabetic population is 1%, then a

similar 50% risk reduction in using drug B, translates into a 0.5% absolute reduction in risk. In this case, the patient and clinician would likely choose drug A given the absolute risk reduction of 5% compared to the alternative absolute risk reduction of 0.5% (5).

For simplicity, the scenario above compares the benefits of two drugs. However, also common that patients and clinicians need to weigh the benefits of an individual drug versus the risks of that same drug. For example, patients considering anticoagulation generally need to weigh the benefit of stroke or embolus prevention against the risk of major bleeding. In these situations, the absolute risk of the medical event trying to be prevented (the benefit) needs to be compared to the absolute risk of an adverse event (the risk). This decision-making exercise will identify patients for whom the benefit of a given treatment far outweigh the risks, as well as patients for whom the risks far outweigh the benefits. If, for simplicity, we assume the absolute risk of adverse events is the same across all patients, then this decision is based purely on the degree of benefit each patient will realize. And if the *relative* benefit is the same for all patients, then their absolute benefit from the treatment is simply a function of their pre-treatment risk (4). For example, an anticoagulant that carries an identical absolute risk of bleeding for all patients, and carries an identical relative risk reduction in stroke for all patients, might only be appropriate for patients who have a very high pre-treatment risk of stroke, as they will see a very large absolute benefit from treatment. For those patients who have a very small pre-treatment risk of stroke, the absolute benefit would be smaller, and might not outweigh the absolute risk. Thus, one of the simplest ways to determine if somebody should receive treatment that is associated with adverse events, is to identify their pre-

treatment risk, in order to estimate the potential absolute benefit of treatment. Such is the rationale behind the CHADS and CHADSVASC2 scores to identify patients with the highest pre-treatment risk, and therefore the greatest potential benefit from anticoagulation. A similar exercise can be done to estimate the pre-treatment risk of an adverse event, and thus to predict the absolute risk of the adverse event occurring with treatment (4).

However, the absolute benefit or risk of a treatment is also a function of relative risk. And relative risk is not necessarily uniform across a population. There may be specific groups of patients who have a particularly high relative risk or a particularly high relative benefit associated with a given treatment. The importance of identifying these different subsets of patients, highlights another limitation of subgroup analyses.

The third challenge facing subgroup analyses the difficulty in defining a subgroup. In some cases, subgroup definitions are clear. For example, it is known that the HER2 receptor plays a key role in the mechanism of action of Trastuzumab, and it would thus make sense to build patient subgroups based on the genetic HER2 characteristic. In such scenarios, in which a subgroup characteristic is clearly a part of a drug's mechanism of action, the use of subgroup analyses are appropriate (4). However, the appropriate definition of a subgroup is not always as clear, and as a result, the usefulness of subgroup analyses is limited. For example, in a study of an anticoagulant, it is difficult to determine whether a subgroup age cutoff should be 65 versus 70 versus 75. It is possible for example, that the particular group of people who are 75 years and older and have

diabetes, have a uniquely great risk reduction from using drug A, or that those who are below 40 years old and do not have diabetes have a particularly strong risk of bleeding. These findings however, would be missed from a clinical trial that made one subgroup based on age greater than 65, and another subgroup based on the presence of diabetes. A similar scenario was seen in the GUSTO trial, which in 1993 found that patients benefited from tPA relative to streptokinase for acute myocardial infarction, and in the subgroup analysis did not find the benefit to be limited to any specific population (6). However, in 2002, Kent et al re-analyzed the GUSTO data using newly developed risk-stratification tools, to predict those patients who would benefit most from treatment, and found that nearly all of the treatment benefits were seen among approximately 50% of patients who were predicted to benefit the most (7). This implied that the approximately 50% of patients with the lower predicted benefits, saw little to no benefit from treatment, and given that they still faced the drug risk of bleeding, should not be given the drug in clinical practice (4).

Given that subgroup analyses are subject to statistical limitations, conflicting relative risk information from different subgroups, and sub-optimal definition of subgroups, a number of authors have proposed greater use of multivariate risk prediction models.

Multivariate Risk Prediction Models

Netflix, one of the most popular users of multivariate risk prediction models, is well-known for its ability to successfully provide individual subscribers with a personalized list of "shows you might like." Netflix could decide which show to recommend to a given subscriber by looking at their entire subscriber base and studying the ratings of "The Crown" versus "Mad Men," and then recommend one of these shows to all subscribers accordingly. Or, they could make a recommendation to a subscriber based on whether or not they were over or under the age of 65, study "The Crown" versus "Mad Men" for these two age based subgroups, and recommend one of the shows accordingly. However, instead of relying on broad studies to predict subscriber preferences, Netflix has developed multivariate risk prediction models, that make use of individualized information captured from every subscriber, such as the genre of shows they watch, and the time of day they watch them, and the speed with which they finish them. With this information, Netflix can develop a multivariate predictive risk model that can classify that subscriber, and precisely predict the which show they will enjoy watching next.

Given the massive amount of data collected from clinical research, genetic and clinical registries, wearable devices, and electronic health records, there is great potential for multivariate risk prediction models in healthcare to identify individual patients who would benefit from specific therapeutics and specific doses. While multivariate models are just one of many strategies to learn from big data, they are an important step towards a world in which a 56 year-old man with diabetes goes into his primary care doctors

office, where the physical activity data from his wearable watch, his daily glucometer readings, his HbA1C from the lab, his other medications from the pharmacy, and information about his other comorbidities, his weight and vital signs and his genetic profile, are all fed into an algorithm that highlights his precise individual risks and benefits from using metformin versus pioglitazone (8).

There should still be a role for "one-size-fits-all" treatment approaches for diseases that are highly prevalent, and have treatments that are highly effective, with minimal risk (9). However, for diseases in which multiple treatment options provide different treatment benefits and different risks of significant adverse events, a precision approach is warranted.

The multivariate risk prediction model has proposed as a method through which the variety of baseline and outcome data, categorical and continuous, can be used to predict patients' individual risks and benefits with a given treatment (4). Multivariate models are generally produced by developing a multivariable linear regression model, and refining the model, by adding or removing variables, in order to strengthen its predictive capabilities (10). Such a model, will produce a risk formula in the format of Y (risk) = $\beta_0 + X_1\beta_1 + X_2\beta_2$ where β_0 is the model's intercept, $\beta_{1\text{-}2}$ are coefficients 1 and 2, and $X_{1\text{-}2}$ are the patients variables. For continuous variables, the value of the variable is entered in the model. For categorical variables, the value of 1 or 0 is entered in the model, depending on whether that variable is or is not applicable to the patient.

To test the theory that multivariate risk models can improve clinical research, and identify a heterogeneity of treatment effect, recent studies have applied new multivariate risk prediction models to completed clinical trials. Salisbury et al assessed the TRITON-TIMI-38 trial which compared prasugrel v. clopidogrel, and developed a multivariate risk prediction model to determine whether there was heterogeneity of treatment effect, not seen in the original trial analysis. They found significant heterogeneity in the both the risks and benefits of clopidogrel and prasugrel for different patients (11). Kernan et al similarly re-analyzed the IRIS trial of pioglitazone after TIA and stroke, and using two risk models developed from the IRIS data, and one external model, also found heterogeneity (12).

Our study applies a similar principle to the RE-LY trial of Warfarin versus Dabigatran for stroke prevention in patients with atrial fibrillation in order to assess the extent to which different patients face different risks and benefits with each treatment.

The RE-LY Trial

Across the United States, atrial fibrillation (AF) affects 2.7-6.1 million people, and its prevalence is expected to increase to 5.6-12 million by 2050 (13). AF is a cardiac arrhythmia, described clinically as an irregularly irregular heart beat due to irregular electrical conduction from the sinoatrial node. This irregular heart rhythm impedes the regular rhythmic output of blood from the heart, and subsequent stasis of blood can lead to development of blood clots in the left atrial appendage. These clots can dislodge, and emboli can enter the cerebral vasculature, and cause a stroke. Due to this pathophysiology, people with AF have a 5 times increased risk for stroke (14).

To prevent stroke, AF is generally controlled with rate control and anticoagulation. Anticoagulation with warfarin, an oral vitamin K antagonist, has historically been the mainstay of treatment, but given the development of Novel Oral Anticoagulants (NOACs), there are now more options for anticoagulation.

While warfarin is an effective treatment for AF, its administration carries substantial lifestyle burdens that can decrease patients' quality of life and decrease adherence. First, patients taking warfarin cannot eat foods that have significant amounts of vitamin K such as spinach or kale, as warfarin's mechanism of action is to antagonize vitamin K. Second, patients who take warfarin need to have their blood INR monitored through regular blood draws, and their warfarin dose adjusted accordingly, in order to ensure a therapeutic PT

and INR, that is sufficient to prevent stroke, but not too high as to increase risk of hemorrhage.

The development of non-vitamin K oral anticoagulants, the anti-X inhibitors apixaban and rivaroxaban and the direct thrombin inhibitor dabigatran, were developed as attractive alternative anticoagulants to warfarin, as they reduce the burden of dietary restrictions and need for frequent blood monitoring. The lifestyle benefits of these medications however, need to be weighed against their clinical effectiveness, and numerous trials have compared the non-vitamin K oral anticoagulants versus warfarin for different clinical conditions (15-17).

Statement of Purpose and Hypothesis

Our study focuses on the RE-LY trial which compared warfarin versus dabigatran in AF for the prevention of stroke (15). The RE-LY trial, conducted in 2009, included 18,113 patients with AF who were deemed to be at increased risk of stroke. This increased risk of stroke was defined by factors such as a history stroke or Transient Ischemic Attack, age or the presence of diabetes mellitus. Exclusion criteria included patients with an increased risk for hemorrhage, a recent or severe stroke and active liver disease (15). Patients were randomized to receive either warfarin, dabigatran 110mg BID or dabigatran 150mg BID. Patients were blinded to their dose of dabigatran, however, were not blinded to receiving warfarin, as blood INR needed to be monitored and warfarin doses adjusted in order to maintain an INR of 2.0 to 3.0. Patients were assessed for two primary outcomes: the presence of stroke or systemic embolism, and the presence of major bleeding. Other outcomes included myocardial infarction, pulmonary embolism and GI bleeding. The study concluded after 5 years, and found the annual rate of stroke to be 1.69% for warfarin, 1.53% for dabigatran 110mg and 1.11% for dabigatran. The annual rate of major bleeding was 3.36% for warfarin, 2.71% for dabigatran 110mg and 3.11% for dabigatran 150mg. The analysis demonstrated that dabigatran 110mg was associated with similar rates of stroke and systemic embolism compared to warfarin, and a significantly lower rate of bleeding. Dabigatran 150mg BID was associated with lower rates of stroke and systemic embolism than warfarin, and a similar rate of bleeding. However, dabigatran 150mg, was associated with a higher rate of gastrointestinal bleeding (15).

While the RE-LY trial subgroup analyses did not reveal significant differences in treatment effect, we sought to determine whether an analysis of the data with multivariate risk models would reveal heterogeneity of treatment effect (15). Using the RE-LY data, we developed two multivariate predictive risk models and assessed individual patients' risks of stroke or systemic embolism and their risk of bleeding, if they were treated warfarin, dabigatran 110mg BID or dabigatran 150mg BID. We then sought to determine whether there were differences individual patients' predicted risks and benefits with each treatment.

The hypothesis of this analysis, is that there is significant heterogeneity of treatment effect with dabigatran versus warfarin, regarding the risk of stroke or systemic embolism, and the risk of major bleeding among the RE-LY population. This hypothesis is tested through the development of a multivariate model, and the subsequent application of that model to the RE-LY data to assess the degree of heterogeneity of treatment among the RE-LY population.

The presence of such a heterogeneity of treatment effect would demonstrate the need to use such a multivariate predictive risk models to aid in day to day clinical decisions regarding warfarin versus different doses of dabigatran for stroke prevention in AF, and would also demonstrate the benefit of using similar models in clinical research going forward.

Methods

Population

The RE-LY trial included 18,113 patients with AF who were at increased risk for stroke (15). Over the 2-year study period, patients were given either warfarin at a titrated dose to achieve an INR of 2.0-3.0, dabigatran 110mg twice a day, or dabigatran 150mg twice a day. Patients were regularly monitored for the presence of the primary outcomes of stroke or systemic embolism or major bleeding, as well as for secondary outcomes including pulmonary embolism, myocardial infarction and gastrointestinal bleeding. To be included in the study, patients were required to have a documented history of AF in the past 6 months, as well as an increased risk of stroke, evidenced by a history of stroke or Transient Ischemic Attack, NYHA class II heart failure symptoms, reduced Left Ventricular Ejection Fraction, age greater than 75, or age 65-74 and the presence of hypertension, coronary artery disease or diabetes mellitus. Patients were excluded from the study if they had an increased risk of hemorrhage, a stroke within the past 15 days, a severe stroke in the past 6 months, a severe heart valve disorder, pregnancy, active liver disease or low creatinine clearance (15). Our study includes all patients who were included in the RE-LY trial and completed a 2-year course of warfarin, dabigatran 110mg BID or dabigatran 150mg. No additional exclusions were applied.

Outcome

The primary outcomes of both the RE-LY trial, and this study are stroke or systemic embolism and major bleeding, representing the major benefits and risks of anticoagulation in patients with AF and a risk of stroke.

Statistical Analyses

Using the data from the RE-LY trial, we developed two multivariate predictive risk models for the two primary outcomes of stroke or systemic embolism and major hemorrhage. Our models were built on the methodological framework developed by Salisbury et al in their analysis of the TRITON-TIMI-38 study data (11). To construct each model, we fit a multivariable linear regression model using all variables that we deemed to be potentially relevant to the outcomes, based on prior research as well as the clinical experience of the authors. We then worked to minimize the number of variables in the model in order to make the models easier to use in the clinical setting. We used the Harrell backwards selection method to the models, to remove variables sequentially, until all variables retained in the model had at least a 5% contribution to the model's predictive capacity (10). The included and excluded variables, and the order in which they were excluded are listed in Table 2, with their corresponding degree of contribution to the model. For each model, we calculated the discrimination (c-statistic) and calibration (Hosmer-Lemeshow, and used restricted cubic splines to assess the assumption that the continuous variables were linearly associated with the outcomes (18).

To assess for heterogeneity of the benefits and risks of each treatment, we applied the risk models to every patient in the RE-LY trial, first assuming they were treated with warfarin, a second time assuming they were treated with dabigatran 110mg, and a third time assuming they were treated with dabigatran 150mg. To characterize the population-level heterogeneity of treatment effect, we determined for each treatment, the mean risk and standard deviation of stroke or systemic embolism and the mean risk of major hemorrhage. To characterize the individual-level heterogeneity of treatment effect, we calculated the absolute difference in risk or benefit between warfarin and dabigatran 110mg, and warfarin and dabigatran 150mg. Specifically, the absolute difference in the risk of stroke or systemic embolism was calculated as each patients' risk if they were treated with dabigatran 110mg minus their risk if they were treated with warfarin, and as their risk if they were treated with dabigatran 150mg minus their risk if they were treated with warfarin. Similarly, the absolute difference in the risk of major hemorrhage is calculated as each individual's risk of major hemorrhage with dabigatran 110mg minus their risk with warfarin, and their risk with dabigatran 150mg minus their risk with warfarin. We report a density plot to demonstrate the range of absolute differences between dabigatran 110mg and warfarin, and dabigatran 150mg and warfarin.

We report categorical variables as frequencies, and continuous variables as medians with interquartile ranges.

All analyses were conducted using R version 3.3.1 and SAS version 9.3.

The initial statistical model was based on the model described by Salisbury et al, and the model development was conducted by Sophie Tang, at Saint Luke's Mid America Heart Institute, with a research group led by Dr. John Spertus (11). Dr. Nihar Desai and I worked closely together to interpret the statistical output for clinical relevance, and worked closely with Sophie Tang to determine the data output most relevant to the hypothesis and project.

For example, I proposed that we create "mock patients" based on the model, and worked with Dr. Desai and Sophie Tang to develop a risk calculator, based on the model output. And using this model, I created the mock patients that are included in the thesis as part of the two decision aids.

Results

A total of 18,040 patients were included in the analysis, 5,983 of whom were assigned to dabigatran 110mg in the RE-LY trial, 6,059 of whom were assigned to dabigatran 150mg, and 5,998 of whom were assigned to warfarin. The median age was 71.4 ± 8.6 and 3.6% of participants were male. The ethnicity of the subjects, was 15.9% Asian, 1.0% Black, 70.0% White, and 13.1% Other. The distribution of AF types of paroxysmal, permanent and persistent were 32.8%, 35.2% and 32.0%, respectively. Baseline characteristics included heart failure (32.0%), hypertension (78.8%), diabetes mellitus (23.3%) and stroke/systemic embolism/TIA (21.8%). As seen in Table 1, demographic and baseline characteristics were similar among all three treatment groups.

Table 1: Baseline Characteristics

	Treatment Pattern Label			Total	P-Value
	Dabigatran 110mg BID n = 5983	Dabigatran 150mg BID n = 6059	Warfarin n = 5998	n = 18040	
Age					0.693
Mean ± SD	71.3 ± 8.6	71.4 ± 8.7	71.5 ± 8.5	71.4 ± 8.6	
Median (IQR)	72.0 (67.0, 77.0)	72.0 (67.0, 78.0)	72.0 (67.0, 77.0)	72.0 (67.0, 77.0)	
Age Grouping					0.121
<40	21 (0.4%)	26 (0.4%)	12 (0.2%)	59 (0.3%)	
40<= and <50	99 (1.7%)	86 (1.4%)	89 (1.5%)	274 (1.5%)	
50<= and <65	873 (14.6%)	916 (15.1%)	849 (14.2%)	2638 (14.6%)	
65<= and <75	2655 (44.4%)	2574 (42.5%)	2635 (43.9%)	7864 (43.6%)	
>=75	2335 (39.0%)	2457 (40.6%)	2413 (40.2%)	7205 (39.9%)	
Sex					0.323
Female	2130 (35.6%)	2228 (36.8%)	2202 (36.7%)	6560 (36.4%)	
Ethnicity					0.871
Asian	948 (15.8%)	961 (15.9%)	955 (15.9%)	2864 (15.9%)	
Black	51 (0.9%)	57 (0.9%)	66 (1.1%)	174 (1.0%)	
White	4191 (70.0%)	4258 (70.3%)	4181 (69.7%)	12630 (70.0%)	
Other	793 (13.3%)	783 (12.9%)	796 (13.3%)	2372 (13.1%)	
Region					0.999
Asia	918 (15.3%)	929 (15.3%)	926 (15.4%)	2773 (15.4%)	
Central Europe	703 (11.7%)	704 (11.6%)	706 (11.8%)	2113 (11.7%)	
Latin America	319 (5.3%)	319 (5.3%)	316 (5.3%)	954 (5.3%)	
USA, Canada	2150 (35.9%)	2195 (36.2%)	2152 (35.9%)	6497 (36.0%)	
Western Europe	1541 (25.8%)	1552 (25.6%)	1543 (25.7%)	4636 (25.7%)	
Other	352 (5.9%)	360 (5.9%)	355 (5.9%)	1067 (5.9%)	
AF Type					0.104
Paroxysmal	1916 (32.0%)	1974 (32.6%)	2030 (33.9%)	5920 (32.8%)	
Permanent	2123 (35.5%)	2183 (36.0%)	2045 (34.1%)	6351 (35.2%)	
Persistent	1941 (32.5%)	1901 (31.4%)	1922 (32.0%)	5764 (32.0%)	
Aspirin at baseline					0.086
	2384 (39.8%)	2338 (38.6%)	2431 (40.5%)	7153 (39.7%)	
CHADS2 Score					0.666
Mean ± SD	2.1 ± 1.1	2.2 ± 1.1	2.1 ± 1.1	2.1 ± 1.1	
Median (IQR)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	
CHADS2 Score					0.221
0	151 (2.5%)	145 (2.4%)	155 (2.6%)	451 (2.5%)	

	Treatment Pattern Label			Total	P-Value
	Dabigatran 110mg BID n = 5983	Dabigatran 150mg BID n = 6059	Warfarin n = 5998	n = 18040	
1	1797 (30.0%)	1810 (29.9%)	1705 (28.4%)	5312 (29.4%)	
2	2081 (34.8%)	2129 (35.1%)	2212 (36.9%)	6422 (35.6%)	
3+	1954 (32.7%)	1975 (32.6%)	1926 (32.1%)	5855 (32.5%)	
History of Heart Failure					0.889
	1929 (32.2%)	1930 (31.9%)	1915 (31.9%)	5774 (32.0%)	
Baseline Heart Failure Classification					0.268
NYHA I	293 (15.2%)	292 (15.1%)	295 (15.4%)	880 (15.3%)	
NYHA II	1222 (63.4%)	1195 (62.0%)	1219 (63.7%)	3636 (63.0%)	
NYHA III	383 (19.9%)	400 (20.7%)	352 (18.4%)	1135 (19.7%)	
NYHA IV	30 (1.6%)	41 (2.1%)	48 (2.5%)	119 (2.1%)	
LVEF					0.723
<=40%	647 (22.0%)	651 (21.9%)	628 (21.2%)	1926 (21.7%)	
Baseline Hypertension Requiring Medical Treatment					0.974
	4711 (78.7%)	4781 (78.9%)	4729 (78.8%)	14221 (78.8%)	
History of Diabetes Mellitus					0.872
	1401 (23.4%)	1398 (23.1%)	1405 (23.4%)	4204 (23.3%)	
History of Stroke/Systemic Embolism/TIA					0.389
	1302 (21.8%)	1357 (22.4%)	1282 (21.4%)	3941 (21.8%)	
Baseline Creatinine Clearance [mL/min]					0.846
Mean ± SD	73.0 ± 27.7	72.8 ± 28.2	73.0 ± 27.4	72.9 ± 27.8	
Median (IQR)	68.7 (53.2, 87.2)	67.9 (53.0, 86.4)	68.5 (53.8, 86.6)	68.4 (53.4, 86.8)	
Creatinine Clearance					0.044
<30	14 (0.2%)	31 (0.5%)	29 (0.5%)	74 (0.4%)	
30<= and <50	1127 (19.7%)	1152 (19.7%)	1048 (18.2%)	3327 (19.2%)	
50<= and <80	2705 (47.2%)	2770 (47.5%)	2794 (48.7%)	8269 (47.8%)	
>=80	1889 (32.9%)	1880 (32.2%)	1872 (32.6%)	5641 (32.6%)	
Weight [kg]					0.385
Mean ± SD	82.9 ± 19.9	82.4 ± 19.3	82.6 ± 19.6	82.7 ± 19.6	
Median (IQR)	80.5 (70.0, 94.0)	80.0 (69.0, 93.0)	80.0 (70.0, 93.0)	80.0 (69.9, 93.4)	
Continuous variables compared using one-way analysis of variance. Categorical variables compared using chi-square or Fisher's exact test.					

As seen in Table 2, patients who took warfarin had the highest rate of stroke (2.1%) and major bleeding (4.7%). Patients who took dabigatran 150mg had the lowest risk of stroke (0.9%), and patients who took dabigatran 110mg had the lowest rate of major bleeding (3.6%).

Table 2: Outcomes

	Treatment Pattern Label			Total	P-Value
	Dabigatran 110mg bid n = 5983	Dabigatran 150mg bid n = 6059	Warfarin n = 5998	n = 18040	
Stroke/Systemic Embolism					< 0.001
	84 (1.4%)	53 (0.9%)	125 (2.1%)	262 (1.5%)	
Major Bleeding					0.011
	216 (3.6%)	261 (4.3%)	281 (4.7%)	758 (4.2%)	

To build the predictive risk models for ischemic stroke or systemic embolism and major bleeding, we used prior literature and clinical expertise, to choose 15 variables that could contribute to the risk of stroke or systemic embolism, and 16 variables that could contribute to the risk of major bleeding. We then conducted two multivariate logistic regressions using all of these 15 variables for stroke and 16 variables for major bleeding. The variables in this model and their analysis of effects are shown in Tables 3-A and 3-B.

Table 3-A: Logistic Full Model: Stroke or Systemic Embolism

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Treatment Pattern	2	9.5803	0.0083
Age	1	0.0413	0.8389
Weight	1	0.8192	0.3654
Sex	1	0.9691	0.3249
Region	4	4.4377	0.3500
Aspirin at Baseline	1	0.0236	0.8778
AF Type	2	2.0821	0.3531
Baseline Heart Failure	1	0.8617	0.3533
Baseline Hypertension	1	1.0227	0.3119
Baseline Diabetes Mellitus	1	13.1876	0.0003
Baseline Stroke/Embolus/TIA	1	28.1507	<.0001
Creatinine Clearance	1	6.3709	0.0116
Age * Treatment Pattern	2	6.7463	0.0343
Baseline Diabetes Mellitus * Treatment Pattern	2	4.4426	0.1085
Creatinine Clearance * Treatment Pattern	2	9.8738	0.0072

Table 3-B: Logistic Full Model: Major Bleeding

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Treatment Pattern	2	27.7766	<.0001
Age	1	1.0372	0.3085
Weight	1	0.7499	0.3865
Sex	1	1.5558	0.2123
Region	4	54.7331	<.0001
Aspirin at Baseline	1	8.3936	0.0038
AF Type	2	3.7635	0.1523
Baseline Heart Failure	1	0.5020	0.4786
Baseline Hypertension	1	1.6975	0.1926
Baseline Diabetes Mellitus	1	11.1931	0.0008
Baseline Stroke/Embolus/ TIA	1	4.9456	0.0262
Creatinine Clearance	1	36.0031	<.0001
Age * Treatment Pattern	2	19.5937	<.0001
Weight * Treatment Pattern	2	14.1518	0.0008
Treatment Pattern * AF Type	4	6.4746	0.1664
Baseline Heart Failure * Treatment Pattern	2	5.1824	0.0749

To create the final predictive risk models, we used the Harrell Backwards Selection Strategy for each model, to sequentially remove the variable that had the smallest contribution to the model's predictive capacity. After the removal of each variable, the model was run again, and the process was repeated until all variables had a p value of less than 0.2. Through this process, 6 variables were removed from the stroke or systemic embolism model (Weight, Sex, Aspirin at Baseline, AF Type, Baseline Heart Failure, Baseline Hypertension) and 2 variables were removed from the major bleeding model (Sex, Baseline Hypertension). Tables 4-A and 4-B show the analysis of effects of the 9 variables included in the final stroke or systemic embolism risk model, and the 14 variables included in the final major bleeding risk model.

Table 4-A: Logistic Prediction Model: Stroke or Systemic Embolism

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Treatment Pattern	2	9.3071	0.0095
Age	1	0.0128	0.9098
Region	4	7.0768	0.1319
Baseline Diabetes Mellitus	1	13.1932	0.0003
Baseline Stroke/Embolus/ TIA	1	27.7175	<.0001
Creatinine Clearance	1	9.8821	0.0017
Age * Treatment Pattern	2	6.5125	0.0385

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Baseline Diabetes Mellitus * Treatment Pattern	2	4.2451	0.1197
Creatinine Clearance * Treatment Pattern	2	9.6253	0.0081

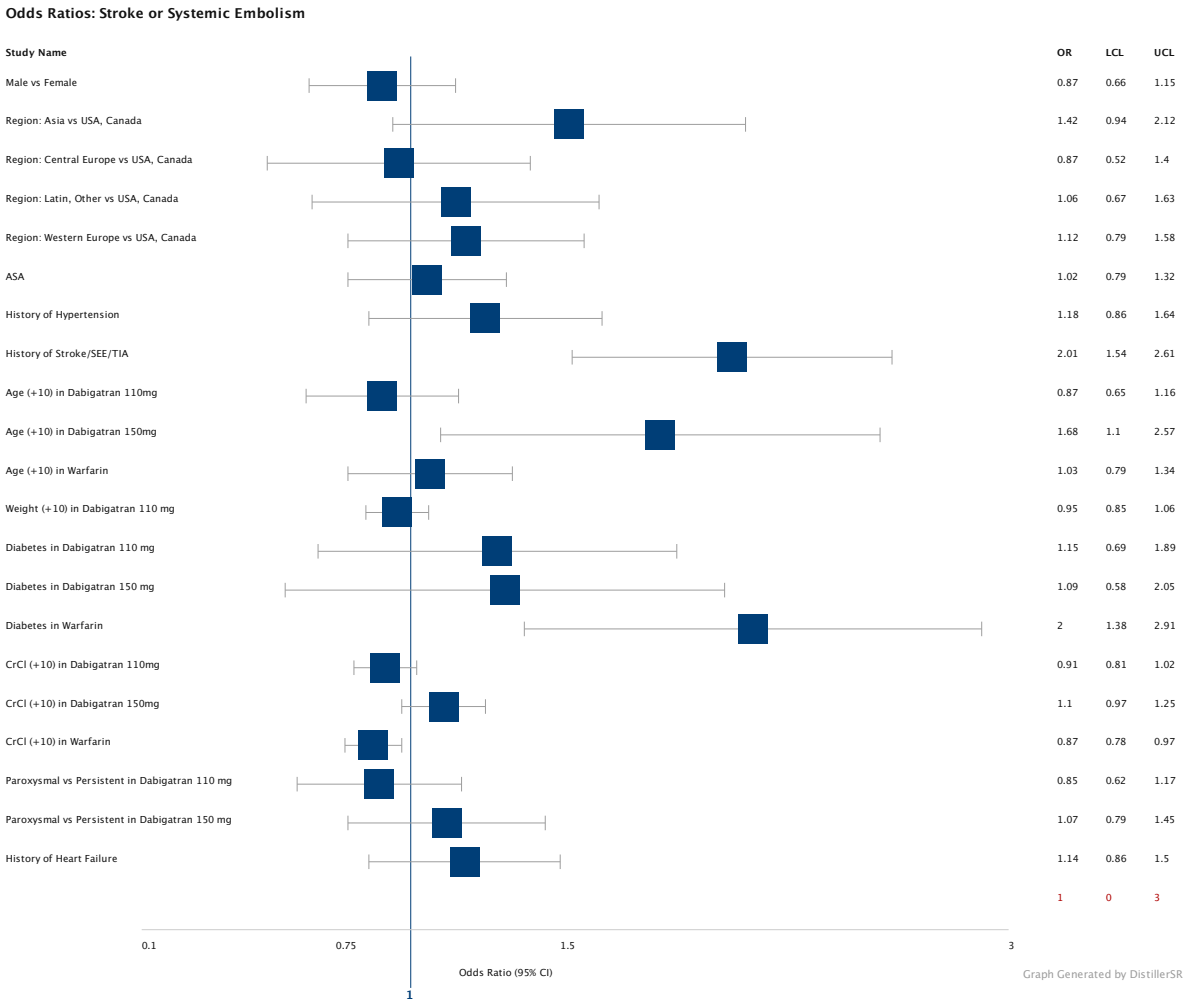
Table 4-B: Logistic Prediction Model: Major Bleeding

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Treatment Pattern	2	27.5394	<.0001
Age	1	1.0161	0.3135
Weight	1	1.6399	0.2003
Region	4	54.7754	<.0001
Aspirin at Baseline	1	8.9961	0.0027
AF Type	2	3.9765	0.1369
Baseline Heart Failure	1	0.4419	0.5062
Baseline Diabetes Mellitus	1	11.7785	0.0006
Baseline Stroke/Embolus/ TIA	1	4.9358	0.0263
Creatinine Clearance	1	37.2981	<.0001
Age * Treatment Pattern	2	19.4564	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Weight * Treatment Pattern	2	14.0595	0.0009
Treatment Pattern * AF Type	4	6.5333	0.1627
Baseline Heart Failure * Treatment Pattern	2	5.3254	0.0698

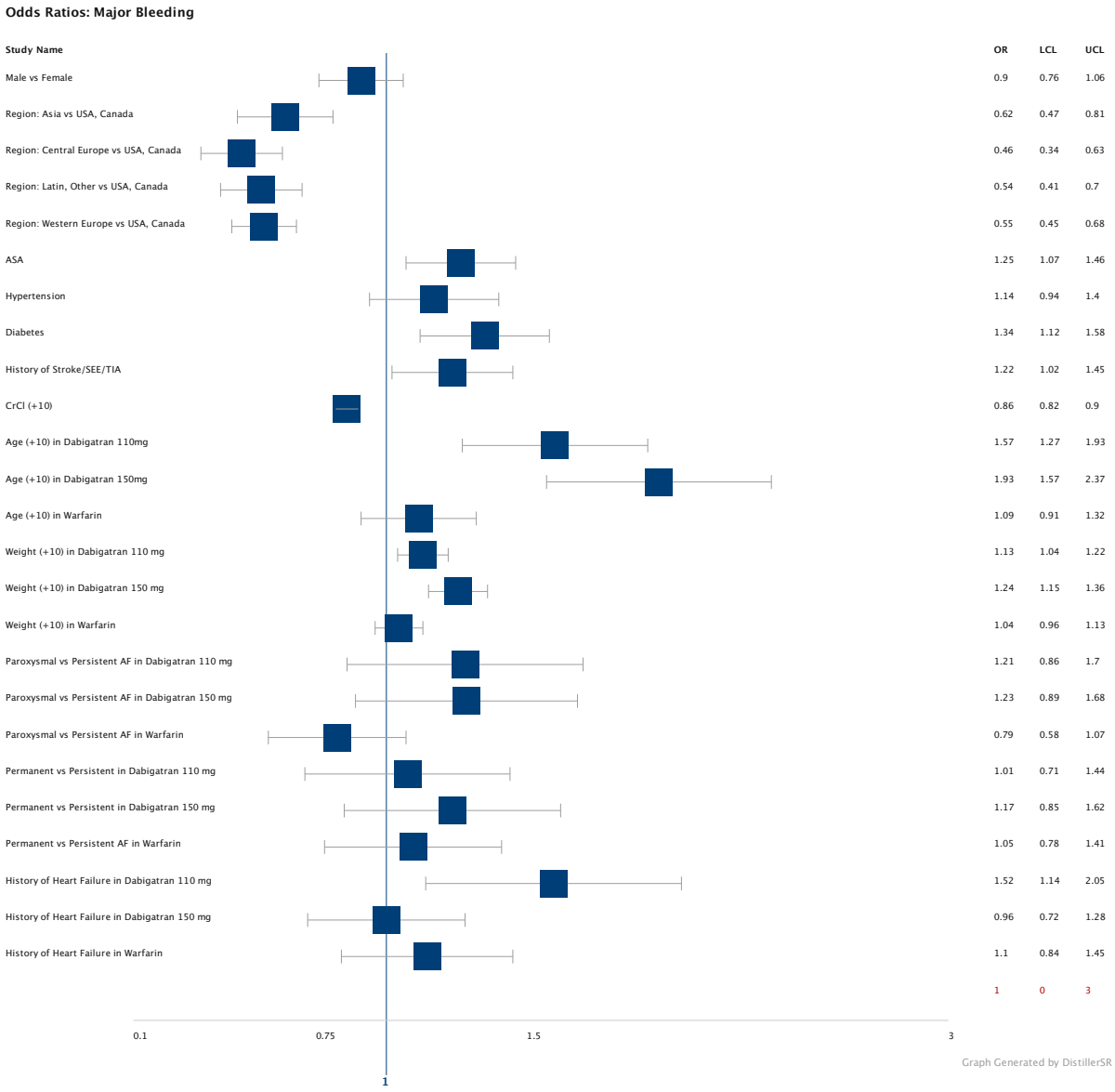
Figures 1-A and 1-B demonstrate the impact of individual patient variables on the full stroke or systemic embolism and major bleeding models respectively. For the stroke or systemic embolism model, a history of stroke/systemic embolism/TIA (OR 2.01) and baseline diabetes with warfarin (OR 2.00) had the largest impact on risk of stroke or systemic embolism. Age in warfarin (OR 1.03), and whether the patient was taking Aspirin (OR 1.02), had the smallest impact on risk of stroke or systemic embolism.

Figure 1-A (19)



For the major bleeding model, history of heart failure in patients taking dabigatran 110mg (OR 1.52), Central Europe versus USA, Canada (OR 0.46), and age in dabigatran 150mg (OR 1.93) had the greatest impact on the risk of bleeding. History of heart failure in dabigatran 150mg (OR 0.96) and weight in warfarin (OR 1.04) had the smallest impact on the model.

Figure 1-B (19)



The final multivariate predictive risk models for stroke or systemic embolism and major bleeding are presented in Tables 5-A and 5-B. The estimates for stroke or systemic embolism show the value of the intercept at -3.3677, and the estimates for major bleeding

show the value of the intercept at -2.8732. The Tables also show the estimates of the regression coefficients for each variable included in the final stroke model or bleeding model. Using these models, the risk of stroke or systemic embolism and the risk of major bleeding can be calculated for any patient based on the equation Y (risk) = $\beta_0 + X_1\beta_{1+} X_2\beta_{2+} X_3\beta_{3+} X_4\beta_4$, where β_0 is the model's intercept, β_{1-4} are the model's regression coefficients, and X_{1-4} are the variables, defined as either 0 or 1 based on the absence or presence of the variable.

Table 5-A: Logistic Prediction Model: Stroke or Systemic Embolism

Analysis of Penalized Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-3.3677	1.2175	7.6507	0.0057
Treatment Pattern	**A**	1	0.6235	1.7072	0.1334	0.7150
Treatment Pattern	**B**	1	-5.9296	2.1931	7.3100	0.0069
Age		1	0.00154	0.0136	0.0128	0.9098
Region	**Asia**	1	0.4461	0.1831	5.9386	0.0148
Region	**Central Europe**	1	-0.0432	0.2414	0.0321	0.8579
Region	**Latin, Other**	1	0.1238	0.2175	0.3241	0.5691
Region	**Western Europe**	1	0.1419	0.1706	0.6924	0.4053
Baseline Diabetes Mellitus		1	0.6902	0.1900	13.1932	0.0003
Baseline Stroke/Embolus/TIA		1	0.6961	0.1322	27.7175	<.0001
Creatinine Clearance		1	-0.0157	0.00498	9.8821	0.0017
Age * Treatment Pattern	**A**	1	-0.0164	0.0196	0.6982	0.4034
Age * Treatment Pattern	**B**	1	0.0495	0.0252	3.8623	0.0494
Baseline Diabetes Mellitus * Treatment Pattern	**A**	1	-0.5543	0.3184	3.0300	0.0817

Analysis of Penalized Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Baseline Diabetes Mellitus * Treatment Pattern	**B**	1	-0.5919	0.3759	2.4803	0.1153
Creatinine Clearance * Treatment Pattern	**A**	1	0.00430	0.00731	0.3451	0.5569
Creatinine Clearance * Treatment Pattern	**B**	1	0.0234	0.00776	9.0839	0.0026

Table 5-B: Logistic Prediction Model: Major Bleeding

Analysis of Penalized Maximum Likelihood Estimates							
Parameter			**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept			1	-2.8732	0.8338	11.8757	0.0006
Treatment Pattern	**A**		1	-3.8326	1.1614	10.8894	0.0010
Treatment Pattern	**B**		1	-5.9535	1.1736	25.7353	<.0001
Age			1	0.00922	0.00915	1.0161	0.3135
Weight			1	0.00504	0.00393	1.6399	0.2003
Region	**Asia**		1	-0.4659	0.1383	11.3387	0.0008
Region	**Central Europe**		1	-0.7622	0.1589	23.0209	<.0001
Region	**Latin, Other**		1	-0.6145	0.1367	20.1987	<.0001
Region	**Western Europe**		1	-0.5979	0.1054	32.1737	<.0001
Aspirin at Baseline			1	0.2334	0.0778	8.9961	0.0027
AF Type	**Paroxysmal**		1	-0.2440	0.1553	2.4680	0.1162
AF Type	**Permanent**		1	0.0521	0.1512	0.1188	0.7303
Baseline Heart Failure			1	0.0919	0.1383	0.4419	0.5062
Baseline Diabetes Mellitus			1	0.2951	0.0860	11.7785	0.0006
Baseline Stroke/Embolus/TIA			1	0.1996	0.0898	4.9358	0.0263

Analysis of Penalized Maximum Likelihood Estimates							
Parameter			**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Creatinine Clearance			1	-0.0152	0.00248	37.2981	<.0001
Age*Treatment Pattern	**A**		1	0.0353	0.0131	7.2561	0.0071
Age*Treatment Pattern	**B**		1	0.0560	0.0130	18.4606	<.0001
Weight * Treatment Pattern	**A**		1	0.00834	0.00495	2.8309	0.0925
Weight * Treatment Pattern	**B**		1	0.0185	0.00493	14.0500	0.0002
Treatment Pattern * AF Type	**A**	**Paroxysmal**	1	0.4333	0.2319	3.4917	0.0617
Treatment Pattern * AF Type	**A**	**Permanent**	1	-0.0351	0.2329	0.0227	0.8802
Treatment Pattern * AF Type	**B**	**Paroxysmal**	1	0.4418	0.2233	3.9148	0.0479
Treatment Pattern * AF Type	**B**	**Permanent**	1	0.1102	0.2217	0.2470	0.6192
Baseline Heart Failure * Treatment Pattern	**A**		1	0.3297	0.2025	2.6516	0.1034
Baseline Heart Failure * Treatment Pattern	**B**		1	-0.1417	0.2012	0.4962	0.4812

The models were tested using a bootstrapping procedure, in which the predictive model was applied to a randomly selected subset of patients, and the model's output, including the predicted risk of stroke, was then compared to the true results of the entire population. This procedure was repeated 10 times, to examine the true predictive capacity of each model. To examine the predicted events with the bootstrapped population compared to the true events of the population, the predicted events of the bootstrapped population were calibrated to the true events of the full RE-LY sample population. Tables 6-A and 6-B show the parameter estimates of the population which, for the slope, are 0.98356 for

the stroke or systemic embolism model and 0.98247 for the major bleeding model, with Pr > |t| values < .001. These results demonstrate that we can reject the null hypothesis that the slope of the model is 0. Tables 7-A and 7-B show the test slope results with Pr>F values of 0.8311 and 0.7318 for the stroke or systemic embolism model and major bleeding model respectively, demonstrating that we cannot reject the null hypothesis that the slope is 1. The proximity of the slope to 1, is demonstrated by the R-square values of 0.9560 for the stroke or systemic embolism model, and 0.9802 for the major bleeding model. The proximity of the slope to 1 demonstrates the strength of the model's predictive capacity. The calibration is demonstrated graphically in Figures 2-A and 2-B, with each event rate for each bootstrap sample against its respective population. The proximity of the intercept towards 0 and the R-Square to 1, and the proximity of the calibration graph to the 45-degree line, demonstrate the linearity of the calibration and therefore the predictive strength of the model. The c-statistic, or area under the curve, is 0.675 for the stroke or systemic embolism model and 0.694 for the major bleeding model.

Table 6-A: Stroke or Systemic Embolism Prediction Model Calibration

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	-0.00022276	0.00136	-0.16	0.8743
Slope	1	0.98356	0.07458	13.19	<.0001

Table 6-B: Major Bleeding Prediction Model Calibration

Table 6-B Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	0.00009327	0.00255	0.04	0.9718
Mean	1	0.98247	0.04938	19.90	<.0001

Table 7-A: Stroke or Systemic Embolism Prediction Model Calibration

Test Slope Results for Stroke or Systemic Embolism				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	2.833789E-7	0.05	0.8311
Denominator	8	0.00000583		
R-Square	0.9560			
C-Statistic	0.675			

Table 7-B: Major Bleeding Prediction Model Calibration

Test Slope Results for Major Bleeding				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.00000251	0.13	0.7318
Denominator	8	0.00001993		
R-Square	0.9802			
C-Statistic	0.694			

Figure 2-A

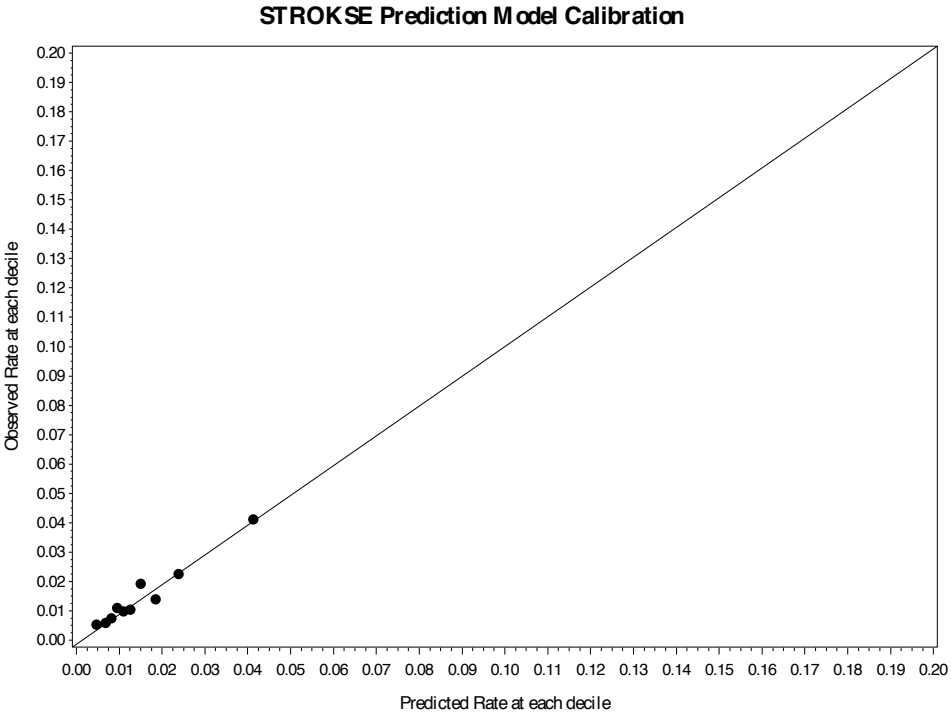
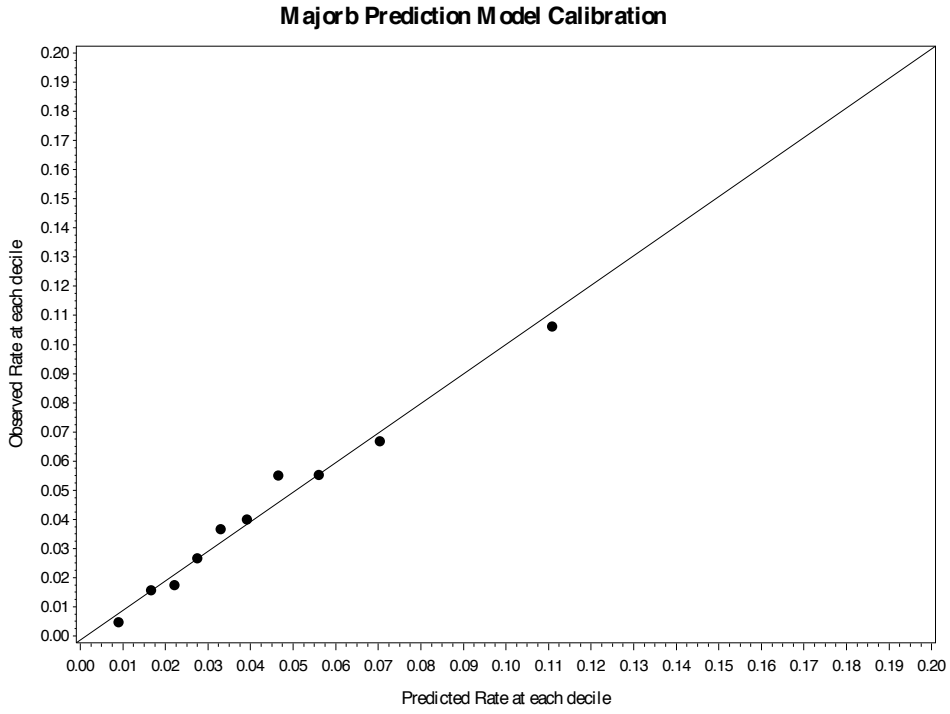


Figure 2-B



We then applied the predictive risk model to each patient in the RE-LY trial to assess for heterogeneity of treatment effect of both dabigatran 110mg and 150mg versus warfarin. The following tables show the number of patients who had a statistically significant difference between their stroke or systemic embolism, or major bleeding risk with dabigatran or warfarin. As seen in Tables 8-A and 8-B, 30.37% and 46.45% of patients had statistically significant differences in their risk of stroke or systemic embolism and major bleeding, respectively, with dabigatran 110mg versus warfarin. As seen in Tables 9-A and 9-B, 70.88% and 47.58% of patients had statistically significant differences in stroke or systemic embolism and major bleeding risk, respectively, with dabigatran 150mg versus Warfarin.

Table 8-A

Stroke/Systemic Embolism log(odds): Dabigatran 110mg - Warfarin Significance				
Significance	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
Non-Significant	12562	69.63	12562	69.63
Significant	5478	30.37	18040	100.00

Table 8-B

Major Bleeding log(odds): Dabigatran 110mg - Warfarin Significance				
Significance	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
Non-Significant	9661	53.55	9661	53.55
Significant	8379	46.45	18040	100.00

Table 9-A

Stroke/Systemic Embolism log(odds): Dabigatran 150mg - Warfarin Significance				
Significance	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
Non-Significant	5254	29.12	5254	29.12
Significant	12786	70.88	18040	100.00

Table 9-B

Major Bleeding log(odds): Dabigatran 150mg - Warfarin Significance				
Significance	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
Non-Significant	9457	52.42	9457	52.42
Significant	8583	47.58	18040	100.00

It is important to note however, that the frequency of patients who had a statistically significantly different risk with dabigatran versus warfarin, includes both patients with an increased risk and patients with a decreased risk. Figures 3 and 4 demonstrate the distribution of this risk, distinguishing patients with and without a statistically significant difference by color and distance from the fit line, and by whether they had an increased or decreased risk, by position above or below the fit line, respectively, with dabigatran. In Figure 3-A for example, all patients who had a statistically significant risk of stroke or systemic embolism with dabigatran 110mg, had a decreased risk. With regard to bleeding risk in Figure 3-B however, while most patients had a decreased risk with dabigatran 110mg, some patients had an increased risk. The greatest distribution between increased

and decreased risk is seen in Figure 4-B, which demonstrates that many patients had a decreased risk of bleeding with dabigatran 150mg, and many patients that had an increased risk of bleeding.

Figures 3A-3B: Dabigatran 110mg - Warfarin Significance

Figure 3-A

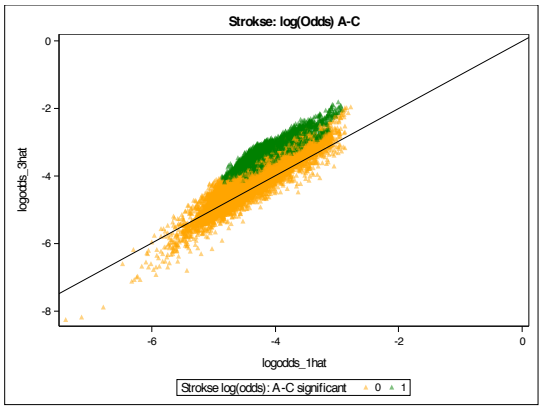
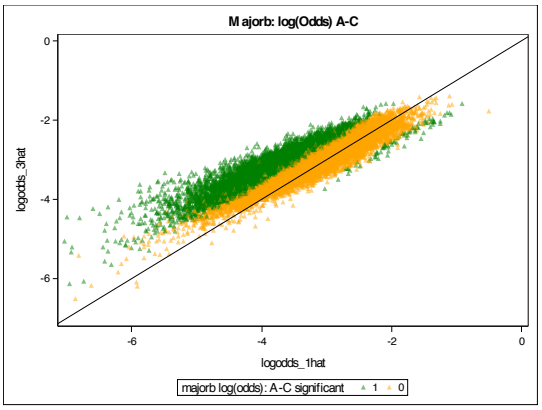


Figure 3-B



Figures 4A-4B: Dabigatran 150mg - Warfarin Significance

Figure 4-A

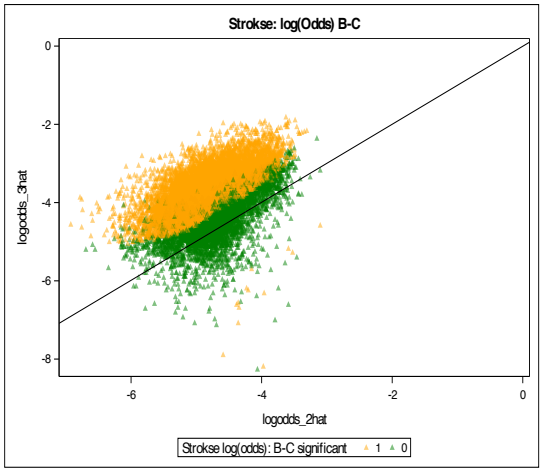
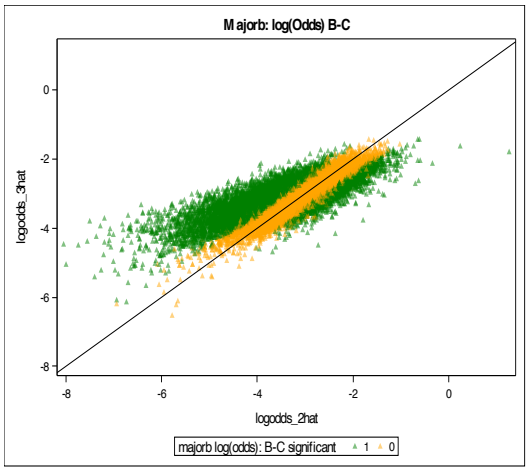


Figure 4-B



To further examine the heterogeneity of treatment effect, we put each patient through the model to determine their risk of stroke or systemic embolism and risk of major bleeding with dabigatran 110mg, dabigatran 150mg, and warfarin. We then calculated the differences in these risks as risk of stroke or systemic embolism with dabigatran minus risk of stroke or systemic embolism with warfarin, and risk of major bleeding with dabigatran minus risk of major bleeding with warfarin. As seen in Table 10, the mean difference of risk between dabigatran 110 and Warfarin was -0.78% ± 0.95% for stroke or systemic embolism and -1.12% ± 1.44% for major bleeding. The mean difference of risk between dabigatran 150mg was -1.32% ± 1.31% for stroke or systemic embolism and -0.41% ± 2.39% for major bleeding. These ranges of risk differences for dabigatran 110mg and 150mg versus warfarin are demonstrated graphically in Figures 5-6. All areas of the curves to the left of 0, represent those patients with a smaller risk of stroke or systemic embolism, or major bleeding with dabigatran, and all areas to the right of the curves represent those patients with a smaller risk of stroke or systemic embolism or major bleeding risk with warfarin. The fact that all curves cross the zero line demonstrates the range of treatment superiority with regard to a given risk. The width of the curves, particularly with regard to major bleeding demonstrates the range of benefits for different patients of using one anticoagulant versus the other.

Table 10: Logistic Predictive Probability Difference between Dabigatran and Warfarin

Logistic Predictive Probability Difference of:	N	Mean	Std Dev	Minimum	Maximum
STROKSE: Dabigatran 110mg - Warfarin	17237	-0.0077530	0.0095157	-0.0932936	0.0215243
Major Bleeding: Dabigatran 110mg - Warfarin	17226	-0.0112073	0.0143663	-0.0776957	0.2324420
STROKSE: Dabigatran 150mg - Warfarin	17237	-0.0131816	0.0131320	-0.1156282	0.0329962
Major Bleeding: Dabigatran 150mg - Warfarin	17226	-0.0041146	0.0238650	-0.0893168	0.6339190

Figure 5

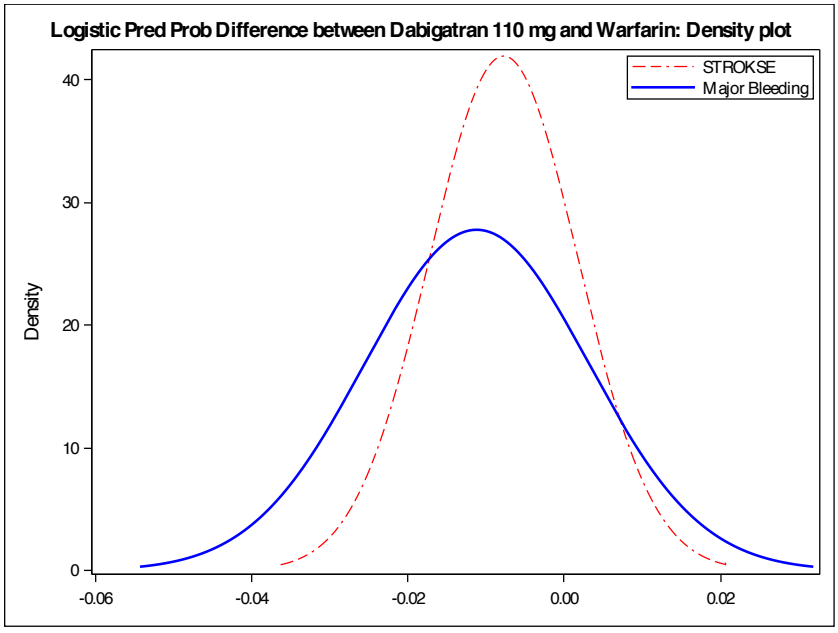
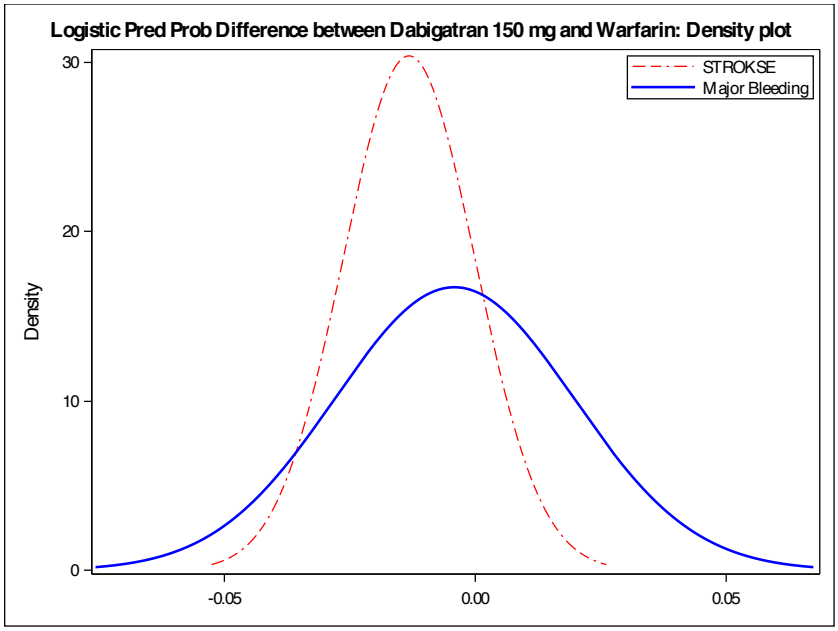


Figure 6



To determine the range of patients for whom dabigatran 110mg or dabigatran 150mg would be superior to warfarin, regarding the risk of stroke or systemic embolism and major bleeding, we plotted the all patients' risk on a chart with a two axes chart, one for each risk (Figures 7 and 8). The X-axis is defined as risk of stroke or systemic embolism with dabigatran minus risk of stroke or systemic embolism with warfarin, and therefore, all patients who fall below the zero line, have a lower risk of stroke or systemic embolism with dabigatran, and those who fall above the zero line have a lower risk of stroke or systemic embolism with warfarin. The Y-axis is defined as risk of major bleeding with dabigatran minus risk of stroke or systemic embolism with warfarin, and therefore all patients who fall to the left of the zero line, have a lower risk of major bleeding with dabigatran, and those who fall to the right of the zero line, have higher risk of major bleeding with dabigatran. In summary, those patients who fall in the lower left quadrant

of the chart would benefit from dabigatran relative to warfarin, as their risk of stroke or systemic embolism and major bleeding are both lower with dabigatran. Those patients who fall in the upper right quadrant of the chart would benefit from warfarin relative to dabigatran, as their risk of stroke or systemic embolism and major bleeding are both lower with warfarin. However, in the upper left and lower right quadrants, there are decisional conflicts. The patients in the upper left-hand corner have a lower risk of stroke with warfarin, but a lower risk of bleeding with dabigatran. The patients in the lower right-hand corner have a lower risk of stroke with dabigatran, but a lower risk of bleeding with warfarin. For those a lower these patients in the upper-left, and lower-right hand corners of the chart, treatment decisions could be made via a well-informed shared-decision-making process, taking into account patients' concerns, values and differences in risk-tolerance between stroke or systemic embolism or major bleeding.

Figures 7 and 8 show that for dabigatran 110mg and 150mg versus Warfarin, there are patients whose stroke or systemic embolism and major bleeding risk combinations fall in all four classifications. The most notable difference between the two figures is that dabigatran 150mg figure has many more patients with a lower risk of bleeding with warfarin, thus explaining why this figure has so many patients who fall in the upper right quadrant where patients have both a lower stroke or systemic embolism and major bleeding risk with warfarin.

Figure 7

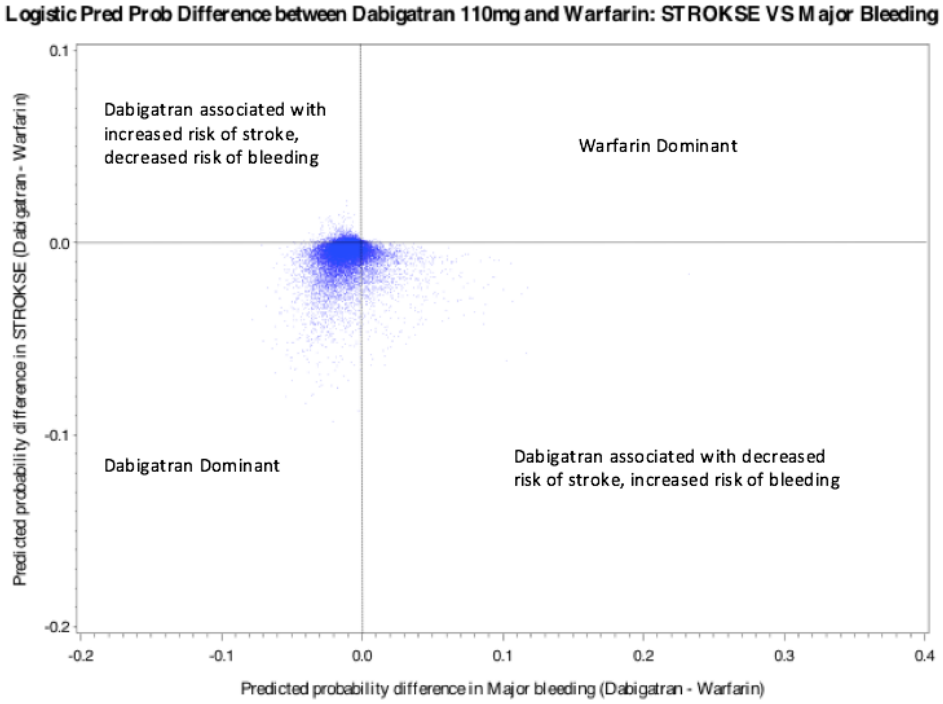
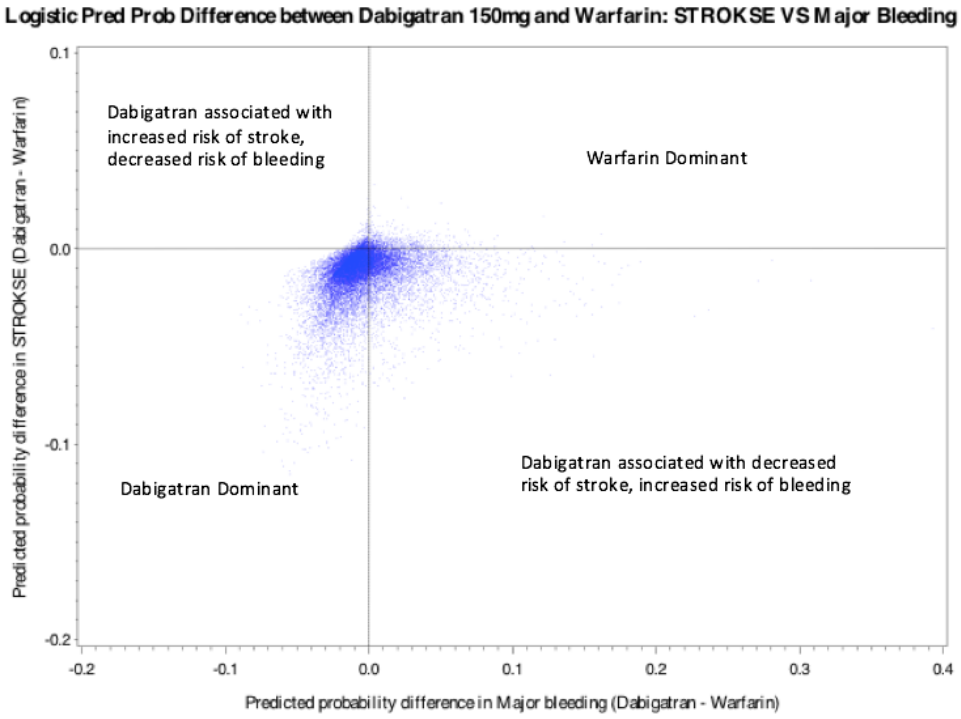


Figure 8



Discussion

In this analysis, we used data from the RE-LY trial to develop two multivariate risk models to predict patients' risk of stroke or systemic embolism, and risk of major bleeding, with treatment with warfarin, dabigatran 110mg and dabigatran 150mg. We applied these risk models to each patient in the RE-LY trial, found significant heterogeneity among patients in the benefits and risks of using the three different medications. For example, some patients derive a much greater benefit than others in stroke or systemic embolism risk reduction from using dabigatran 110mg versus warfarin. And other patients derive a much greater risk than others in major hemorrhage when using dabigatran 150mg compared to warfarin. The range of this heterogeneity of treatment effect is demonstrated by Table 8 and figures 6-7, which show the distribution of the differences in absolute risk reduction for each patient when using dabigatran 150mg and dabigatran 110mg versus warfarin. The mean difference (dabigatran minus warfarin) in the risk of stroke was -0.78% ± 0.95% for dabigatran 110mg and -1.32% ± 1.31% for dabigatran 150mg. The mean difference in the risk of major bleeding was -1.12% ± 1.44% for dabigatran 110mg and -0.41% ± 2.39% for dabigatran 150mg. These findings confirm our hypothesis that there is a heterogeneity of treatment effect with regard to dabigatran and warfarin.

The heterogeneity of treatment effect was explained by Figures 2-A and 2-B, which show the contributions of individual characteristics to the risk of stroke or major bleeding. Certain variables such as a history of diabetes or heart failure increased the risk of stroke

or systemic embolism or major bleeding, respectively. Whereas other variables, such as male gender or region decreased the risk of stroke or systemic embolism or major bleeding, respectively. These opposing risk factors, with their different weight, can complicate clinical decision making, and highlight the need for multivariate prediction models to accurately and objectively weigh individual risks and benefits.

These findings build on recent studies to demonstrate the importance of analyzing clinical trials for heterogeneity of treatment effect using multivariate risk models. Salisbury et demonstrate the heterogeneity of treatment effect between clopidogrel and prasugrel in the TRITON-TIMI 38 trial, and by Kernan et al to demonstrate the heterogeneity of treatment of pioglitazone post stroke or TIA in the IRIS trial (11,12). Our application of multivariate risk models to the RE-LY trial similarly found heterogeneity of treatment effect, that was previously unknown, adding to further evidence regarding the potential benefit of multivariate predictive risk models to clinical care.

To assess the potential impact of these findings, it is important to return to the current state of clinical practice and decision making with regard to dabigatran and warfarin. In the U.S. clinicians and patients choose between doses of dabigatran 150mg and 75mg and warfarin based on the findings of clinical trials such as RE-LY, and their respective subgroup analyses. However, while the findings of these trials report means, medians, standard deviations and the results that are best for the population as whole, they don't necessarily reflect the results for patients at the extremes of the treatment effect. For example, the RE-LY trial found that for the population as a whole, there were similar

risks of major bleeding with dabigatran 150mg versus warfarin warfarin (15). Our analysis does not dispute this finding, but rather notes that for certain patients there is a significantly increased risk of bleeding with dabigatran 150mg compared to warfarin.

Trials such as RE-LY do publish subgroup analyses to account for this heterogeneity of treatment effect. However, these subgroup analyses have numerous statistical limitations, can pose opposing risk contributions from the application of different subgroups, and are not easy for a clinician to practically apply to individual patients in the clinical setting. Furthermore, while these subgroup analyses certainly narrow the population to the individual, they still generalize results to a population that shares only one characteristics. This is akin to Netflix recommending a movie to a 70-year-old woman, because that movie has been popular among people over the age of 65. Much like Netflix is able to make a movie prediction to this woman based on her age, gender, the time it takes her to finish a TV show, and the actresses in previous shows she has finished, modern medicine has the potential to make individualized medication recommendations based on multiple individual characteristics.

Most importantly, our findings that there are both patients for whom dabigatran or warfarin would reduce their risk of stroke and bleeding, demonstrates the potential for more individualized decisions to improve patient health. These data are displayed graphically in figures 7 and 8 showing each patients risk along two axes, for risk of stroke and risk of major bleeding. Those patients who are in the lower left, and upper right quadrants showed both a greater absolute risk reduction in both stroke or systemic

embolism and major hemorrhage with dabigatran and warfarin, respectively. The challenge in the clinical setting, is to identify which quadrant a given patient would fall in, such that their risk of stroke or major bleeding could be minimized.

The basis of these findings in a multivariate predictive risk model, highlights the potential for this model to be used in clinical practice to predict patients risks of stroke or bleeding with dabigatran versus warfarin. Before this risk model is used clinically, it will need to be tested in further populations. However, if validated, it's use in clinical practice would be beneficial, and feasible. Multivariate predictive risk models are already in widespread use, with clinicians and patients using multivariate risk scores such as CHADSVASC2 and HASBLED to predict their individual benefits and risk with using anticoagulation. This multivariate predictive risk model could be used in a similar way to help clinicians and patients make a fully informed, individualized decision about a medication at the point of care.

While such a tool would identify those patients who would clearly benefit from warfarin or dabigatran with regard to both stroke and bleeding risk, there will also be patients who have the lowest stroke risk with one medication, and the lowest bleeding risk with the other medication. These patients are represented by the upper left and lower right quadrants in Figures 7 and 8. These treatment decisions pose a challenge, that will require careful evaluation of the degree of risks and benefits with different given treatment, and a shared decision-making process to identify the treatment that would best align with patient preferences and values.

To facilitate the decision-making around treatments that carry both a benefit and a risk with regard to other available treatments, Salisbury et al, proposed the calculation of a "net clinical benefit score" (11). The net clinical benefit is defined as the benefit to risk ratio. As applied to our findings, net clinical benefit would be defined as the ratio of the reduction of risk in stroke or systemic embolism (the benefit), to the increase in risk of bleeding. For each patient, the benefit of a given treatment, or reduction in risk of stroke or systemic embolism, would be calculated as the absolute difference of the predicted risk of stroke or systemic embolism for warfarin minus the predicted risk with dabigatran. The risk, or increased risk of major hemorrhage, would be calculated as the absolute difference of the predicted risk of major hemorrhage with warfarin minus the predicted risk with dabigatran. For each patient a benefit to risk ratio >1, or an absolute benefit with dabigatran that is greater than the absolute risk with dabigatran, would signify a net benefit with dabigatran relative to warfarin. A ratio of <1, or an absolute benefit with dabigatran less than the absolute risk increase with dabigatran, indicates a net benefit with warfarin relative to dabigatran.

In clinical practice, patients and clinicians may have their own individual perspectives on the amount of benefit required to outweigh a given risk. For example, some patients and clinicians might feel that preventing stroke or systemic embolism is twice as important as avoiding major hemorrhage, while others might feel that preventing stroke or systemic embolism is half as important as avoiding major hemorrhage. In the first group who put greater emphasis on reducing stroke, a benefit to risk ratio of >0.5 would be required to signify a net clinical benefit with dabigatran relative to warfarin. For the second group

who put greater emphasis on avoiding hemorrhage, a benefit to risk ratio of >2 would be required to signify a net clinical benefit with dabigatran relative to warfarin (11).

The presentation of such a net clinical benefit score could help patients and clinicians simplify challenging decisions. Such a tool could be part of a larger strategy to communicate risk in an intuitive fashion, and to facilitate the use of multivariate predictive models in clinical decision making.

Effective presentation of risk models to facilitate shared decision making

The promise of individualized patient data, is its potential to guide clinical decision-making. Thus, as we develop methods to provide more individualized data, we must also develop tools to present the data in a way that can most effectively aid decision-makers.

These tools can be used as part of the shared-decision making model, which is designed to enhance patient clinician communication, and ensure treatment decisions that are backed by all available evidence, and align with patient values (20). Shared decision making is recommended by the American Heart Association for anticoagulant treatment decisions in AF, to ensure that treatment decisions are aligned with patient's values (14).

The shared decision-making model consists of a clinician and patient forming a partnership, through which the clinician explains the risks and benefits of different treatment options, patients explain their experiences, values and thoughts about different treatments, and together the patients and clinicians discuss their thoughts about the treatment options, and agree on a course forward (20-23). There is of course a range of the degree to which different patients wish to be involved in the decision-making process, and it is important for clinicians, at the outset, to ask patients what they would prefer. This might range from the patient seeking information and making the decision, to the patient asking the doctor's opinion and then making the decision, to asking the doctor to make a decision by themselves (24). However, it is important that clinicians do not assume patient deference to their provider, as it has been shown that patients randomized

to participate in shared decision making versus usual care, report greater satisfaction (20,25).

Visual decision aids can greatly enhance the shared decision-making process by helping the clinician convey treatment risk and benefit information in a clear and easy to understand manner (20). Effective decision-aids incorporate a number of principles that have been learned from behavioral decision-making science and studies of past decision aids.

First, decision-aids should avoid use of number needed to treat statistics and relative risk statistics. Number needed to treat is a difficult concept for many patients to understand, and relative risk can lead both patients and clinicians to perceive an exaggerated risk or treatment effect. As an example of this phenomenon with relative risk, a patient who is told that a given treatment will reduce their risk of stroke by 50%, is will likely be more inclined to use the drug than if they were told that it reduced their risk of stroke from 2% to 1% (26). Furthermore, as previously discussed, a relative risk reduction of 50% may have significantly different implications for a patient who sees their risk decrease from 20% to 10% versus a patient who sees their risk decrease from 2% to 1%. Therefore, in order to avoid an exaggeration of treatment effect, and to ensure that risks are truly individualized, decision-aids should present risks as absolute risks (24,27,28).

In addition to presenting risks as absolute risks, it is also important for decision aids to convey the incremental change in risk, from the patient's baseline risk. If a clinician tells

a patient that they have a 5% risk of major bleeding over the next 10 years with a given anticoagulant, they may incorrectly assume that all of their bleeding risk is due to the anticoagulant, and that their risk without treatment would be 0%. Therefore, if the patient has a 4% baseline risk of bleeding without treatment over the course of 10 years, it is important to convey that information, and to explain that taking the anticoagulant will add 1% to their absolute risk of bleeding over 10 years (27,28).

Similarly, it may be helpful to provide patients with contextual absolute risk information about their other health risks, as a basis of comparison for the risks associated with treatments. For example, if a patient is weighing whether to take a drug that is associated with a 1% absolute risk of bleeding over 10 years, it may be helpful for them to understand that their absolute risk of carotid artery disease over that same period is 40% (27).

When comparing the absolute risks associated with multiple treatments, it is important for decision aids to convey those risks in consistent formats. That is, if one risk is presented as a percent, the other risk should be presented as a percent, not as a frequency ratio. People, especially with lower numeracy, which tends to decrease with age, may have difficulty comparing a 7% risk to a 5/100 risk, and it has been found that people tend to be biased, in perceiving frequencies as being greater than percents (27,29,30). Therefore, a 7% risk should be compared to a 5% risk, or a 7/100 risk should be compared to 5/100 risk. Furthermore, if the risks are presented as frequency ratios, it is important that the ratios have identical denominators. A 2/50 risk of bleeding with drug A

should not be compared to a 6/75 risk with drug B. Rather, this should be presented as a 4/100 risk with drug A, compared to an 8/100 risk with drug B. Additionally, it is best to avoid presenting frequencies in a "1 in x" format, as this has been found to be difficult for patients to understand (27,31). And, when possible, it is best to use smaller denominators. A 1/10 risk is found to be much better intuitively understood than a 1/100 risk (27,32).

There is significant debate on the role of narrative language on decision aids. The generally agreed upon notion is that "words matter." The challenge then is to determine when they serve to add helpful additional information, versus when they introduce or exacerbate anecdote bias. For example, the inclusion of patient testimonials on decision aids has been found to make patients more concerned about the severity of bad outcomes (27,33). For a patient who views a cerebral hemorrhage as akin to a bruise on their forehead, the inclusion of such a testimonial would be warranted. However, for the patient whose friend recently died of a cerebral hemorrhage, the sharing a testimonial of cerebral hemorrhage might further bias them to overestimate the likelihood of a hemorrhage in their decision-making process. In fact, one of the greatest benefits of visual decision-aids, is their ability to reduce anecdote bias, and allow patients and clinicians to consider data is as objective a way as possible, so they can make decisions in line with their values (27,34).

Language can also be added to decision-aids in order to "label" results. For example, rather than only telling patients that the risk of hemorrhage is 1% with one drug and 5% with another drug, a decision aid could label the 1% risk as "low," and the 5% risk as

"high." Patients given decision aids with labels were found to be more likely to incorporate risk information in their decision-making (27,35).

Decision-aids can introduce "framing effects" through which different presentations of identical information, can lead to different patient decisions. Numerous studies have demonstrated that patients are more likely to tolerate risky medications or procedures if the risks of procedures were framed in a positive as opposed to negative frame (27,36). For example, Levin et al, showed that people were hypothetically more willing to undergo a procedure with a 50% success rate than a procedure with a 50% failure rate (37). This line of reasoning suggests that patients would be more inclined to take an anticoagulant if they are told "95% of patients with AF who take Warfarin don't have a stroke within 10 years," compared to being told that "5% of patients with AF who take warfarin have a stroke within 10 years."[b] The presence of framing effects should not be used to influence patient decisions, but should be considered as potential sources of biases in patient decisions.

The "recency effect" is another bias that can be introduced by decision aids. In this bias, patients are more likely to place greater emphasis on the piece of information they heard last (28,38). For example, if a patient is first told about an anticoagulant's association with hemorrhage, and are then told about its' effect on stroke prevention, they may place disproportionate weight on the drug's impact on stroke prevention. And if they are first told about stroke prevention, and then told about hemorrhage, they may place greater

[b] This a hypothetical deduction of previous study findings, and has not itself been studied.

emphasis on hemorrhage.[c] To reduce this bias, decision-aids can summarize all presented information, prior to the patient decision, such that the value patients place on a given treatment characteristic, is consistent with their own personal values and not with the order in which it was presented (28).

The visual presentation of decision-aids may also influence comprehension. First, while the use of numbers to present statistical information, may be the best way to convey precise, or verbatim, information, there are many benefits to conveying information through pictographs. Pictographs present statistics by using different colored icons (usually ovals, smiley faces, or bathroom symbols) to represent the affected proportion of an at-risk population. For example, to convey that 3/100 people who take an anticoagulant get a cerebral hemorrhage within 10 years, a pictograph might display 100 bathroom figures, and color three of them red, signifying the likelihood of cerebral hemorrhage. Though it has been proposed that these three red figures be spread out among the pictograph, to emphasize the randomness of an event occurring, it has been found that comprehension increases when they are grouped together (27,39). Pictographs can be intuitively easy to understand, and have been shown to be better than bar graphs in conveying "gist" knowledge. Furthermore, while the use of numbers is the most effective way to convey precise, or verbatim, information, pictographs are superior to pie graphs in conveying verbatim knowledge (28,40).

The major benefit of pictographs however, is their ability to address "denominator neglect" (41). Through this bias, people tend to overweigh numerators, and under weigh

[c] This a hypothetical deduction of previous study findings, and has not itself been studied.

denominators. For example, many people are likely to believe that a 15/100 risk is greater than a 2/10 risk (41). Icon arrays draw people's eyes to the denominator, to help them understand the context of the numerator and to get a more accurate understanding of their risk. The presentation of too much information can distract patients from key points such as the denominator, and it is therefore essential that pictographs include only the most important information (28,42).

In addition to "neglecting" the denominator, it is also common for people to misunderstand the meaning of the denominator. For example, a patient who is told they have a 3/100 risk of hemorrhage, may assume that every 3/100 times they take an anticoagulant, they can expect to get a hemorrhage, and that nearly everybody eventually gets a hemorrhage (27,43). Icon arrays provide a shared mental model through which clinicians can work with patients to ensure that they accurately understand the meaning of the denominator and how it relates to their individual risk.

Finally, icon arrays have the potential to address neglect of a time course. Risk is defined as events/at risk population over a given time period. However, this time period is often neglected. For example, patients who are told they have a 10-year, 3/100 risk of hemorrhage may assume that every single year, for 10 years, 3 people out of 100 will suffer a hemorrhage, for a total of 30 people out of 100 at the end of 10 years. Icon arrays should clearly display the time course of a risk, and again, help clinicians and patients work together under a shared mental model (27).

The use of shared decision making with decision aids increases patients' knowledge of pre-treatment risk and treatment options, increases patient engagement and decreases decisional conflict (20,44,45). Furthermore, decision-aids that are tailored to patient's individual profile have been shown to have a particularly strong positive impact on patient comprehension (27,46). There have also been number of studies of shared decision making and decision aids regarding anticoagulation in AF. While many of these studies were done without individual risk models and prior to non-vitamin K oral anticoagulants, similar to other studies of shared decision-making, they found that patients who underwent shared decision making, had greater understanding of treatment options, and decreased decisional conflict (20,47).

The multivariate predictive risk models for stroke or systemic embolism and major bleeding developed in this analysis, if validated, could be incorporated into decision-aids in order to help clinicians and patients with AF and an increased risk of stroke, work together to make decisions about anticoagulation with dabigatran versus Warfarin. The following icon arrays were generated using http://www.iconarray.com, a tool developed by the Risk Science Center at the University of Michigan, which incorporates many of the lessons from the decision-aid literature, to create effective decision-aids (48).

These are examples of decision-aids that could be presented to two different patients. The first patient is Frank, a 62-year-old, 85 kg man with diabetes, a creatinine clearance of 85 ml/min, a history of TIA, and permanent AF, who is taking Aspirin, and has no history of heart failure. The second patient is Maria, an 87-year-old, 58 kg woman who also has

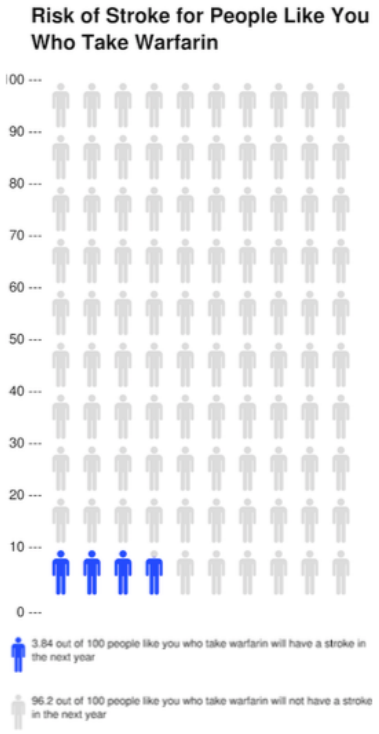
diabetes, a creatinine clearance of 85ml/min, has persistent AF and is taking aspirin, and has no history of stroke, systemic embolism, TIA or heart failure.

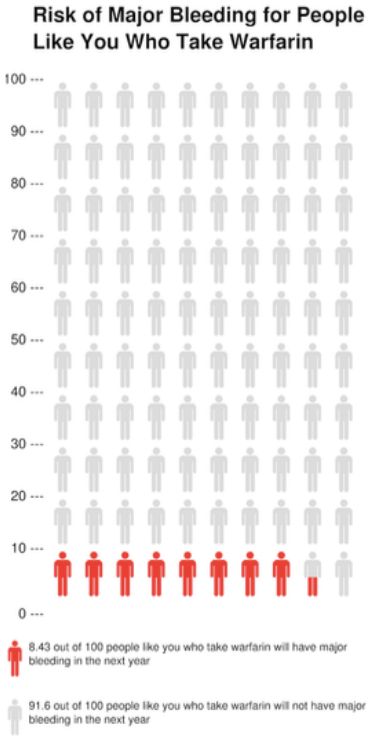
For each patient, the decision-aids show their risk of stroke with dabigatran 150mg and warfarin, side-by side, and below that, their risk of major bleeding with dabigatran 150mg and warfarin, side-by side.

As seen in the decision-aids, both patients face decisional challenges. Frank's risk of stroke is higher with warfarin, at 3.84% than it is with dabigatran, at 0.92%. However, his risk of major bleeding is lower with warfarin, at 8.43% than it is with dabigatran, at 13.1%. Maria's risk of stroke is slightly higher, with warfarin, at 2.4% than it is with dabigatran, at 1.49%. However, her risk of major bleeding is much lower with warfarin, at 8.51% than it is with dabigatran at 22.12%.

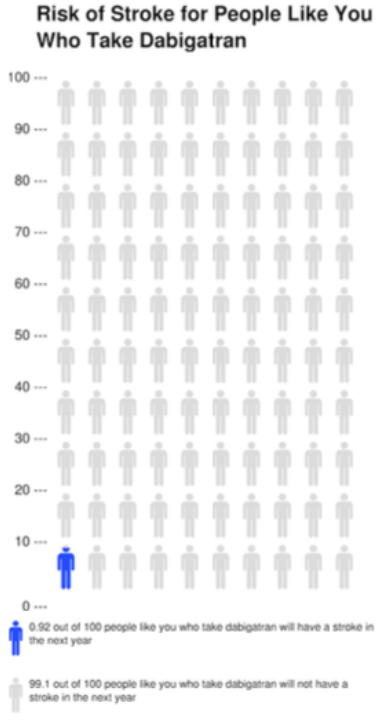
Decision-Aid for Frank, a 62 year-old, 85 kg man from USA, with diabetes, a creatinine clearance of 85 ml/min, a history of TIA, and permanent AF, on Aspirin, with no history of heart failure

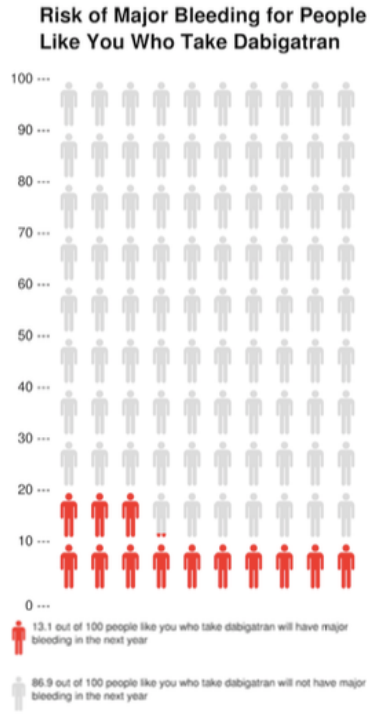
Warfarin





Dabigatran

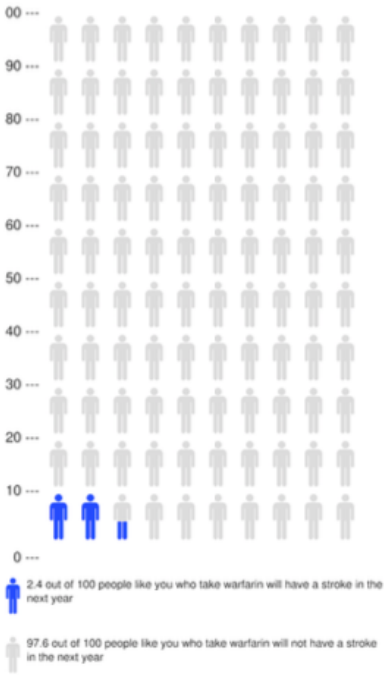




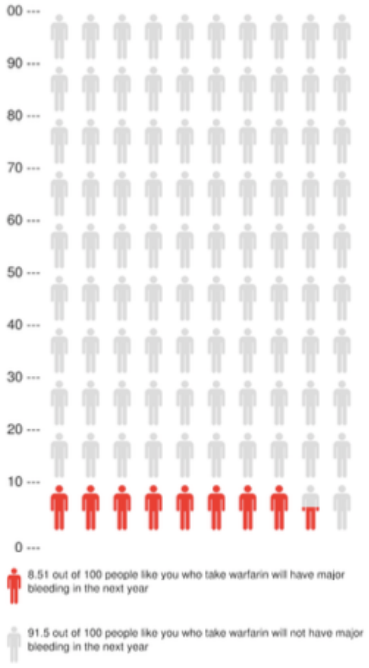
Decision-Aid for Maria, an 87 year-old, 58 kg woman from the USA, with diabetes, a creatinine clearance of 74 ml/min, and persistent AF, on Aspirin, with no history of stroke/SEE/TIA or heart failure

Warfarin

Risk of Stroke for People Like You Who Take Warfarin

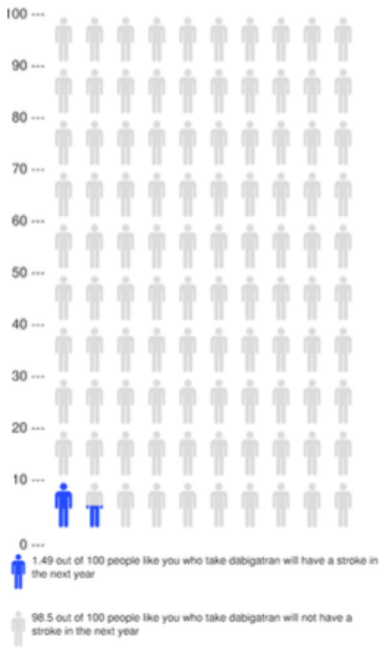


Risk of Major Bleeding for People Like You Who Take Warfarin

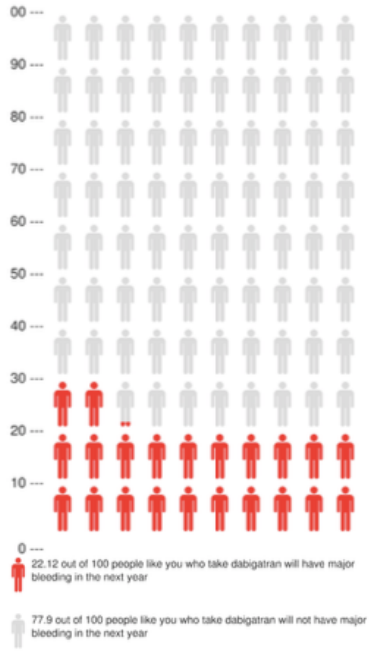


Dabigatran

Risk of Stroke for People Like You Who Take Dabigatran



Risk of Major Bleeding for People Like You Who Take Dabigatran



Frank and Maria, could both look at these personalized decision-aids with their physician, as part of a regular office visit, in order to understand the different risks and benefits associated with an anticoagulant, and alongside other considerations such as lifestyle implications, could make a fully informed choice that aligns with their respective values. Frank might have a very active lifestyle, and given his relatively young age of 62 and his history of TIA, might be very concerned about a stroke causing functional limitation, and might be willing to accept the risk of major bleeding requiring a transfusion. In this case, Frank would likely choose dabigatran given that it decreases his risk of stroke by close to 3% relative to warfarin, even though it is associated with an over 4% increased risk of bleeding. Maria on the other hand, at age 87, might have numerous functional limitations, and might be more concerned about being hospitalized for major bleeding issues than she is about new functional limitations from a stroke. In this case, Maria might be unimpressed by the approximately 1% decreased risk of stroke with dabigatran relative to warfarin, but might be very concerned by the nearly 14% increased risk of major bleeding with dabigatran and would therefore be likely to choose warfarin. In such cases where patients place different weight on the risks of stroke and major bleeding, a net clinical benefit score as discussed earlier, could also play a role in calculating for the patient, which treatment choice would best align with their values and preferences.

Such an aid could also theoretically be interactive, and further engage patients by providing them an opportunity change the variables in the model. For example, Frank could see how his risk of stroke will change as he gets older, and Maria could see how her risk of major bleeding would change if her creatinine clearance were to decrease.

It is important to note that this graphic has not been tested, and only serves to highlight how a pictograph decision-aid could potentially convey information from the predictive risk model to patients as part of their usual care. Decision-aids should be studied for effectiveness with regard to decisional factors such as patient comprehension, decisional conflict and decision comfort.

While risk models can be created, and decision-aids can be built, they will only be valuable if patients and clinicians use them during clinical practice. Given the potential for numerous risk models to be developed and updated over the coming years and decades, it is important to build an infrastructure that provides patients and clinicians with the latest and most relevant risk models in an easily accessible manner. One potential example of such an infrastructure, would be a risk calculator that is automatically built in to the electronic medical record, and that quickly displays decision aids onto a tablet that is readily available in the exam room, or to a patient's smartphone. In addition to improving awareness and accessibility of newly developed risk models, one of the biggest challenges to the use of shared decision making and decision aids, is their potential impact on clinical workflow. Studies have found that shared decision making can add approximately 3 minutes to a clinical encounter (20,49). While incorporating automated systems into electronic medical records may add some efficiency, it is also important that shared decision making is valued by payers and health systems, such that time for shared decision making is allocated as a part of standard practice.

Conclusion

This study demonstrates that there is heterogeneity of treatment effect between dabigatran 110mg and dabigatran 150mg versus Warfarin for the RE-LY population. These findings were demonstrated through a multivariate predictive risk model that has the potential to be used in clinical practice to predict treatment effect. However, prior to use in clinical practice, the effectiveness predictive risk model will need to be evaluated against other populations. If validated, this model could enhance the shared decision-making process by providing patients with easy-to-understand individualized information about their predicted risks and benefits with a given treatment.

However, there are a number of limitations to our study.

First, the multivariate predictive risk model has not been tested against independent populations outside of the RE-LY trial. To account for this limitation, we conducted a bootstrapping procedure, which tested the model on small samples of the RE-LY population against the rest of the population. However, internally validated models, due to the fact that they developed and tested with the same population, may be subject to limitations in generalizability. Thus, prior to use in clinical practice, it is important for this model to be tested external, independent samples.

Second, the generalizability of our model is limited by the inclusion and exclusion criteria of the RE-LY trial. Patients in the RE-LY trial were limited to those with AF and

an increased risk of stroke. Patients who were pregnant, had active liver disease, a creatinine clearance below 30 ml/min, or had a severe heart valve disorder were excluded from the study. However, the RE-LY trial population was designed to include those patients most likely to face a treatment decision of dabigatran versus warfarin, and despite the limitations of the inclusion and exclusion criteria, the model designed in this study should still have widespread clinical applicability.

Third, the c-statistic, or area under the curve of our model, was only is 0.675 for the stroke model and 0.694 for the major bleeding model. These c-statistics represent some limitation in the degree of discrimination of the model and its ability to predict events versus non-events. However, as noted by Salisbury et al, predictive models with modest c-statistics are superior to the generalization of findings across a broad population, and models with c-statistics greater than 0.60 have been shown to be effective (4,11).

Last, the models in this study only predict individualized risk for stroke and systemic embolism, and major bleeding. Patients may also be interested in other individualized information, such as their risk of myocardial infarction with each treatment option, or lifestyle implications of each medication. However, stroke and systemic embolism and major bleeding were selected for this study, because they are generally seen as the biggest risk and benefit concerns of anticoagulant choice in patients with AF and an increased risk of stroke. And these individualized risks do not need to be used in isolation, but rather as information alongside other considerations such as lifestyle implications as part of a comprehensive treatment decision.

Despite these limitations, this analysis shows that heterogeneity of treatment effect exists in the RE-LY trial, building on evidence that it exists in other trials. The multivariate model developed in this study demonstrates the potential of this model, if further validated, to be used in clinical practice to help patients understand their individualized risk and benefit with regard to taking warfarin or different doses of dabigatran.

The increasing amount of data available in medicine poses a potential for unique statistical profiles to be developed for patients, such that they can make individualized decisions to maximize their benefits, minimize their risks and improve their health. The use of multivariate predictive risk models, such as that published in this analysis, have the potential to move clinical practice closer to this goal of precision medicine.

But for multivariate predictive risk models to be used effectively in clinical practice, their findings must be clearly communicated and easily understood. In addition to recognizing the potential of unique statistical profiles to help batters hit a baseball, Ted Williams communicated this potentially complicated concept through a simple illustration of the strike zone that could be intuitively understood by boys and girls across cultures and generations. Using the science behind visual decision aids and effective presentation of statistics, we too can work to ensure that it is just as intuitive for a patient and clinician to use a risk model to choose between dabigatran and warfarin, as it was for a young baseball hitter to use Ted Williams' strike zone graphic to choose "a good ball to hit" (1).

References

1. Williams T, Underwood, J., Cupp, R. The Science of Hitting. New York, NY: Simon & Schuster, Inc.; 1970.
2. FACT SHEET: President Obama's Precision Medicine Initiative. https://obamawhitehouse.archives.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative: The White House, Office of the Press Secretary; 2015.
3. Collins FS, Varmus H. A new initiative on precision medicine. The New England journal of medicine 2015;372:793-5.
4. Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. BMC medical research methodology 2006;6:18.
5. Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. BMJ (Clinical research ed) 2015;351:h5651.
6. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The New England journal of medicine 1993;329:673-82.
7. Kent DM, Hayward RA, Griffith JL, et al. An independently derived and validated predictive model for selecting patients with myocardial infarction who are likely to benefit from tissue plasminogen activator compared with streptokinase. Am J Med 2002;113:104-11.
8. Krumholz HM. Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System. Health Affairs 2014;33:1163-70.
9. Mega JL, Sabatine MS, Antman EM. Population and personalized medicine in the modern era. Jama 2014;312:1969-70.
10. Harrell FE. Regression Modeling Strategies. Secaucus, NJ: Springer-Verlag New York, Inc; 2006.
11. Salisbury AC, Wang K, Cohen DJ, Li Y, Jones PG, Spertus JA. Selecting Antiplatelet Therapy at the Time of Percutaneous Intervention for an Acute Coronary Syndrome. Weighing the Benefits and Risks of Prasugrel Versus Clopidogrel 2013;6:27-34.
12. Kernan WN, Viscoli CM, Dearborn JL, et al. Targeting Pioglitazone Hydrochloride Therapy After Stroke or Transient Ischemic Attack According to Pretreatment Risk for Stroke or Myocardial Infarction. JAMA neurology 2017;74:1319-27.
13. Go AS, Mozaffarian D, Roger VL, et al. Heart Disease and Stroke Statistics—2014 Update. A Report From the American Heart Association 2014;129:e28-e292.
14. January CT, Wann LS, Alpert JS, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. Circulation 2014;130:e199-267.
15. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. The New England journal of medicine 2009;361:1139-51.

16. Hijazi Z, Siegbahn A, Andersson U, et al. High-sensitivity troponin I for risk assessment in patients with atrial fibrillation: insights from the Apixaban for Reduction in Stroke and other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial. Circulation 2014;129:625-34.
17. Larsen TB, Skjoth F, Nielsen PB, Kjaeldgaard JN, Lip GY. Comparative effectiveness and safety of non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study. BMJ (Clinical research ed) 2016;353:i3189.
18. Hosmer DW, Lemeshow, S., . Applied Logistic Regression. New York: John Wiley and Sons, Inc; 2000.
19. DistillerSR Forest Plot Generator from Evidence Partners. 2018. at https://www.evidencepartners.com/resources/forest-plot-generator/ - forest_plot_12_graph_edit_linebyline.)
20. Seaburg L, Hess EP, Coylewright M, Ting HH, McLeod CJ, Montori VM. Shared decision making in atrial fibrillation: where we are and where we should be going. Circulation 2014;129:704-10.
21. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. The New England journal of medicine 2012;366:780-1.
22. Godolphin W. Shared decision-making. Healthcare quarterly (Toronto, Ont) 2009;12 Spec No Patient:e186-90.
23. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Social science & medicine (1982) 1999;49:651-61.
24. Lin GA, Fagerlin A. Shared Decision Making. State of the Science 2014;7:328-34.
25. Weymiller AJ, Montori VM, Jones LA, et al. Helping patients with type 2 diabetes mellitus make treatment decisions: Statin choice randomized trial. Archives of Internal Medicine 2007;167:1076-82.
26. Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. Medical decision making : an international journal of the Society for Medical Decision Making 2007;27:638-54.
27. Trevena LJ, Zikmund-Fisher BJ, Edwards A, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Medical Informatics and Decision Making 2013;13:S7.
28. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. Journal of the National Cancer Institute 2011;103:1436-43.
29. Peters E, Hart PS, Fraenkel L. Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. Medical decision making : an international journal of the Society for Medical Decision Making 2011;31:432-6.
30. Peters E, Hess TM, Vastfjall D, Auman C. Adult Age Differences in Dual Information Processes: Implications for the Role of Affective and Deliberative Processes in Older Adults' Decision Making. Perspectives on psychological science : a journal of the Association for Psychological Science 2007;2:1-23.

31. Pighin S, Savadori L, Barilli E, Cremonesi L, Ferrari M, Bonnefon JF. The 1-in-X effect on the subjective assessment of medical probabilities. Medical decision making : an international journal of the Society for Medical Decision Making 2011;31:721-9.
32. Garcia-Retamero R, Galesic M. Using plausible group sizes to communicate information about medical risks. Patient education and counseling 2011;84:245-50.
33. Betsch C, Ulshofer C, Renkewitz F, Betsch T. The influence of narrative v. statistical information on perceiving vaccination risks. Medical decision making : an international journal of the Society for Medical Decision Making 2011;31:742-53.
34. Fagerlin A, Wang C, Ubel PA. Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? Medical decision making : an international journal of the Society for Medical Decision Making 2005;25:398-405.
35. Peters E, Dieckmann NF, Vastfjall D, Mertz CK, Slovic P, Hibbard JH. Bringing meaning to numbers: the impact of evaluative categories on decisions. Journal of experimental psychology Applied 2009;15:213-27.
36. Moxey A, O'Connell D, McGettigan P, Henry D. Describing treatment effects to patients. Journal of general internal medicine 2003;18:948-59.
37. Levin IP, Schnittjer SK, Thee SL. Information framing effects in social and personal decisions. Journal of Experimental Social Psychology 1988;24:520-9.
38. Ubel PA, Smith DM, Zikmund-Fisher BJ, et al. Testing whether decision aids introduce cognitive biases: results of a randomized trial. Patient education and counseling 2010;80:158-63.
39. Ancker JS, Weber EU, Kukafka R. Effect of arrangement of stick figures on estimates of proportion in risk graphics. Medical decision making : an international journal of the Society for Medical Decision Making 2011;31:143-50.
40. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. Patient education and counseling 2008;73:448-55.
41. Garcia-Retamero R, Galesic M, Gigerenzer G. Do Icon Arrays Help Reduce Denominator Neglect? Medical Decision Making 2010;30:672-84.
42. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. A demonstration of ''less can be more'' in risk graphics. Medical decision making : an international journal of the Society for Medical Decision Making 2010;30:661-71.
43. Gigerenzer G, Galesic M. Why do single event probabilities confuse patients? BMJ (Clinical research ed) 2012;344:e245.
44. Hess EP, Knoedler MA, Shah ND, et al. The Chest Pain Choice Decision Aid. A Randomized Trial 2012;5:251-9.
45. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database of Systematic Reviews 2017.
46. Albada A, Ausems MG, Bensing JM, van Dulmen S. Tailored information about cancer risk and screening: a systematic review. Patient education and counseling 2009;77:155-71.

47. Thomson RG, Eccles MP, Steen IN, et al. A patient decision aid to support shared decision-making on anti-thrombotic treatment of patients with atrial fibrillation: randomised controlled trial. Quality & safety in health care 2007;16:216-23.
48. Images created by Iconarray.com. Risk Science Center and Center for Bioethics and Social Sciences in Medicine, University of Michigan. Accessed 2018-01-03.
49. Montori VM, Shah ND, Pencille LJ, et al. Use of a Decision Aid to Improve Treatment Decisions in Osteoporosis: The Osteoporosis Choice Randomized Trial. The American Journal of Medicine 2011;124:549-56.