Journal of the American Heart Association

ORIGINAL RESEARCH

Impact of a Patient's Baseline Risk on the Relative Benefit and Harm of a Preventive Treatment Strategy: Applying Trial Results in Clinical Decision Making

Tamar I. de Vries, MD, PhD; Manon C. Stam-Slob, MD, PhD; Ron J. G. Peters, MD, PhD; Yolanda van der Graaf, MD, PhD; Jan Westerink, MD, PhD; Frank L. J. Visseren, MD, PhD

BACKGROUND: For translating an overall trial result into an individual patient's expected absolute treatment effect, differences in relative treatment effect between patients need to be taken into account. The aim of this study was to evaluate whether relative treatment effects of medication in 2 large contemporary trials are influenced by multivariable baseline risk of an individual patient.

METHODS AND RESULTS: In 9361 patients from SPRINT (Systolic Blood Pressure Intervention Trial), risk of major adverse cardiovascular events was assessed using a newly derived risk model. In 18 133 patients from the RE-LY (Randomized Evaluation of Long-Term Anticoagulant Therapy) trial, risk of stroke or systemic embolism and major bleeding was assessed using the Global Anticoagulant Registry in the Field–Atrial Fibrillation risk model. Heterogeneity of trial treatment effect was assessed using Cox models of trial allocation, model linear predictor, and their interaction. There was no significant interaction between baseline risk and relative treatment effect from intensive blood pressure lowering in SPRINT (P=0.92) or from dabigatran compared with warfarin for stroke or systemic embolism in the RE-LY trial (P=0.71). There was significant interaction between baseline risk and treatment effect from dabigatran versus warfarin in the RE-LY trial (P<0.001) for major bleeding. Quartile-specific hazard ratios for bleeding ranged from 0.40 (95% CI, 0.26–0.61) to 1.04 (95% CI, 0.83–1.03) for dabigatran, 110 mg, and from 0.61 (95% CI, 0.42–0.88) to 1.20 (95% CI, 0.97–1.50) for dabigatran, 150 mg, compared with warfarin.

CONCLUSIONS: Effect modification of relative treatment effect by individual baseline event risk should be assessed systematically in randomized clinical trials using multivariate risk prediction, not only in terms of treatment efficacy but also for important treatment harms, as a prespecified analysis.

REGISTRATION: URL: https://www.clinicaltrials.gov; Unique identifier: NCT01206062.

Key Words: adverse drug events ■ cardiovascular disease ■ multivariate analysis ■ thromboembolism ■ treatment outcome

Every patient is different, and every patient will react differently to medication. However, randomized clinical trials (RCTs) usually report results as a single relative effect size. In current clinical practice, this single relative treatment effect is then used and applied to diverse patient categories and a plethora of individual patients. Implicitly, the assumption is made that this single relative effect measure is true for all study participants, independent of an individual's characteristics.

The anticipated *absolute* treatment effect is then dependent on not only the relative treatment effect,

Correspondence to: Frank L. J. Visseren, MD, PhD, University Medical Center Utrecht, PO Box 85500, F02.224, 3508 GA Utrecht, the Netherlands. E-mail: F.L.J.Visseren@umcutrecht.nl

Supplementary material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.120.017605

For Sources of Funding and Disclosures, see page 9.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- For translating an overall trial result into an individual patient's expected absolute treatment effect, differences in relative treatment effect between patients need to be taken into account, preferably assessed with a multivariable risk score.
- Individual patient data from large clinical trials provide a unique opportunity to assess relative treatment effect modification across the full range of baseline risk (risk without the allocated intervention), as determined by multiple characteristics.

What Are the Clinical Implications?

- Individual absolute treatment effects should be estimated for making treatment decisions in clinical practice, not only in terms of treatment efficacy but also for important treatment harms.

Nonstandard Abbreviations and Acronyms

GARFIELD-AF	Global Anticoagulant Registry in the Field–Atrial Fibrillation
RE-LY	Randomized Evaluation of Long-Term Anticoagulant Therapy
SE	systemic embolism
SPRINT	Systolic Blood Pressure Intervention Trial

but also on baseline risk: for example, a relative risk reduction of 20% for an intervention will result in a 5% absolute risk reduction for a patient with a 25% baseline 10-year risk for the outcome, and a 2% absolute risk reduction for a patient with a 10% baseline 10-year risk. However, for these absolute treatment effects to be true, the assumption of a single relative treatment effect independent of baseline risk and different clinical characteristics needs to hold. An individual patient in clinical practice is not the same as the average trial participant. The dilemma in clinical practice is whether a single overall relative treatment effect is also true for the patient a health care professional is seeing. The best treatment *on average* may not be the best treatment for a given patient. Moreover, a given patient may experience more important treatment *harms* than another patient.

Simple relative treatment effect modification is regularly assessed in trials using subgroup analyses, a one-characteristic-at-a-time approach. This article explains why such approaches to subgroup analyses could have limitations, and suggests an alternative method to evaluate relative treatment effect modification. In this article, we evaluate whether relative treatment effects of medication in 2 large contemporary trials are influenced by the baseline risk of an individual patient. This method can be used to assess treatment effect heterogeneity (ie, how the relative treatment effect varies across patients) in terms of both treatment benefit and treatment harm.

Subgroup Analyses: One at a Time?

Subgroup analyses, based on single patient characteristics, are frequently performed to assess differences in the relative treatment effect between groups of patients. However, there are major limitations to subgroup analyses in RCTs.[1–4] First, stratification and subsequent estimation of relative treatment effects within numerous subgroups results in a high risk of chance findings.[1,3] Second, one-at-a-time subgroup analyses introduce a "reference class problem."[1] For example, if both age and sex are effect modifiers, which relative treatment effect measure is the "correct" one for a young woman? Furthermore, by selecting subgroups on >1 variable at a time (eg, by making subgroups based on *both* age and sex), a low number of end points in each subgroup would preclude reliable subgroup analyses.

A formal statistical interaction test between a patient characteristic of interest and the treatment allocation in a trial is more accurate than the estimation of subgroup-specific treatment effects. However, a true interaction may not be detected as few trials are adequately powered for single variable-treatment interaction analysis. At the same time, if there is no actual interaction effect, the probability of finding a false-positive treatment interaction is 5% per tested characteristic, of which there are many. Therefore, if effect modification by a single factor *can* be expected, this should be prespecified in the trial design and taken into account in trial power estimation.

Assessment of Relative Treatment Effects by Baseline Risk

Baseline risk, the risk without the trial intervention, for a clinical outcome can be estimated using a risk model composed of multiple prognostic factors, either based on an existing risk model or derived in the trial itself. Assessment of relative treatment effect modification by individual baseline risk can be used as a method to assess treatment interactions. If no treatment effect modification is expected on the basis of previous data or biological mechanisms, a multivariable approach to relative treatment effect modification has important advantages over one-at-a-time subgroup analyses. First, as this method

does not rely on stratification into subgroups, sufficient power may be maintained to assess treatment effect differences in the study population. Second, it is possible that a combination of patient characteristics, rather than a single patient characteristic at a time, influences the treatment effect from an intervention. It is also possible that a patient characteristic not assessed in subgroup analysis contributes to relative treatment heterogeneity. As many patient characteristics are correlated, it may be possible that even if this variable is not included in the risk model, heterogeneity based on baseline risk is found. Third, a single multivariable test for treatment effect modification by baseline risk prevents chance findings attributable to multiple testing.[1] Furthermore, this approach is based on the data from the trial itself, can be published with the main results from the trial, and, if relative treatment effect heterogeneity is apparent, this information facilitates clinical decision making. Finally, relative treatment effect heterogeneity not only pertains to efficacy but also to safety. Therefore, although no clear heterogeneity may be present on treatment efficacy in a trial, there may still be clinical relevant heterogeneity in treatment safety, with important consequences for individualized clinical decision making.

Statistical Analysis

The first step is to estimate baseline risk for individual patients, preferably with an existing risk score. If not all variables are collected in the studies, a new risk score needs to be derived. A time-to-event survival model (eg, a Cox proportional hazards model) is subsequently fitted with each participant's linear predictor of the risk model using their patient characteristics (which determines an individual's baseline risk) and treatment allocation as predictors. Potential relative treatment effect modification by baseline risk is assessed by adding the interaction term "treatment*linear predictor" to a Cox model. As an example, the following notation is used in R (where LP denotes the model linear predictor, and allocation the trial allocation):

```
> cph(Surv(follow.up,endpoint)~LP * allocation, data=data, x=T, y=T)
```

The models with and without interaction term are compared using the likelihood ratio test with a $P<0.05$ indicating a significant interaction. If a statistically significant effect modification is found, the next question is whether this statistical significant difference in relative treatment effect is clinically relevant.

By stratifying the study population into quartiles and estimating the relative treatment effect within these quartiles (quartile-specific hazard ratio [HR] or relative risk reduction) and comparing these, the presence of relative treatment effect heterogeneity may be assessed. The stratification of the study population into quartiles is merely a way to study the change in relative treatment effect with differing baseline risk, and quartile-specific HRs cannot be used in clinical practice.

METHODS

Study Populations

Data were used from the SPRINT (Systolic Blood Pressure Intervention Trial) and the RE-LY (Randomized Evaluation of Long-Term Anticoagulant Therapy) trial. Because of the sensitive nature of the data collected for this study, requests to access the SPRINT data set from qualified researchers trained in human subject confidentiality protocols may be sent to the National Heart, Lung, and Blood Institute Biologic Specimen and Data Repository Information Coordinating Center at https://biolincc.nhlbi.nih.gov/. Requests to access the RE-LY trial data set may be sent to www.clinicalstudydatarequest.com.

SPRINT (registration number: NCT01206062) included 9361 patients aged ≥50 years with systolic blood pressure of ≥130 mm Hg and an increased risk for cardiovascular disease. Eligible participants were randomized 1:1 to an intensive target systolic blood pressure of <120 mm Hg, or a standard target systolic blood pressure of <140 mm Hg. The study was ended prematurely on the basis of a reduced risk in the primary composite major adverse cardiovascular event end point at interim analysis. The RE-LY trial (registration number: NCT00262600) included 18 133 patients with atrial fibrillation, and randomized them 1:1:1 to warfarin, dabigatran, 110 mg twice daily, or dabigatran, 150 mg twice daily. Most participants had an indication for oral anticoagulation therapy based on their Score for Atrial Fibrillation Stroke Risk. The RCT demonstrated that in patients with atrial fibrillation, dabigatran given at a dose of 110 mg twice daily was associated with a similar risk of stroke and systemic embolism (SE) compared with warfarin, while having lower rates of major bleeding. Dabigatran at a dose of 150 mg twice daily was associated with lower rates of stroke and SE and similar rates of major bleeding compared with warfarin. Detailed descriptions of both trials and inclusion and exclusion criteria have been published previously.[5–8] (Ethical) approval was obtained from the national regulatory authorities and ethical committees of the participating centers, and all participants provided written informed consent.

Statistical Analysis

Individual Treatment Effect Estimation of Intensive Blood Pressure Lowering

Because existing risk scores, like the Framingham risk score, Systematic COronary Risk Evaluation risk chart, or atherosclerotic cardiovascular disease risk score, could not be used for risk estimation because

of missing variables and different definitions of the primary outcome,[9–11] a prediction model for the primary composite end point (myocardial infarction, other acute coronary syndromes, stroke, heart failure, or death from cardiovascular causes) was derived in the control arm of SPRINT using Cox proportional hazards analysis. Well-known predictors were selected from previously published risk scores.[9–13] Details on model development and model validation for estimation of the risk of the primary end point are presented in the appendix (Data S1 and Figure S1). First, baseline mortality risk was estimated by entering individual patient characteristics in the model formula. Using the method as described above, heterogeneity of treatment effect was then assessed using the linear predictor from the newly derived risk model.

Individual Treatment Effect Estimation of Dabigatran on Stroke and Major Bleeding in Atrial Fibrillation

Baseline 1-year risk for the primary end point stroke and SE and for the risk of major bleeding was estimated using the externally validated Global Anticoagulant Registry in the Field–Atrial Fibrillation (GARFIELD-AF) risk models. Details on the model derivation and end point definitions have been published previously and are summarized in the appendix (Data S2).[14] Model performance was assessed with the c-statistic (95% CI) for discrimination and with calibration plots of predicted versus observed risk (Figure S2). First, baseline risks of stroke/SE and major bleeding were estimated by filling in individual patient characteristics in the GARFIELD-AF model formulas. Using the method described above, heterogeneity of treatment effect was then assessed using the linear predictors of the stroke/SE and major bleeding GARFIELD-AF risk functions.

RESULTS

Baseline Characteristics and Events

Baseline characteristics of both trials are shown in the Table. During a median follow-up of 3.2 years (interquartile range, 2.7–3.8 years) in SPRINT, 319 end points occurred (6.8%) in those who received standard treatment, compared with 243 end points (5.2%) in patients who received intensive treatment. During a median follow-up of 2.0 years in the RE-LY trial (interquartile range, 1.6–2.4 years), 160 primary events (2.7%) and 426 major hemorrhages (7.1%) occurred in participants who received warfarin, compared with 295 primary events (2.4%) and 757 major hemorrhages (6.3%) in patients who received dabigatran in either dose.

Table. Baseline Characteristics of SPRINT and RE-LY Trial Study Populations

Characteristics	SPRINT (n=9361)	RE-LY trial (n=18 113)
Age, y	68±9	71±9
Male sex	6029 (64)	11 514 (64)
Race or ethnicity		
White	5399 (58)	12 616 (70)
Black	2802 (30)	176 (1)
Hispanic	984 (11)	879 (5)
Other*	176 (2)	4442 (24)
Current smoking	1244 (13)	5979 (33)
History of cardiovascular disease	1562 (17)	5248 (29)
Systolic blood pressure, mm Hg	140±16	131±18
Body mass index, kg/m^2	30±6	29±6
Total cholesterol, mmol/L	4.9±1.1	4.7±1.1
LDL cholesterol, mmol/L	2.9±0.9	N/A
HDL cholesterol, mmol/L	1.4±0.4	N/A
Estimated glomerular filtration rate, mL/min per 1.73 m^2	67±24	73±28
Uses a statin	4083 (44)	8057 (45)
Uses antihypertensives	8479 (91)	14 509 (80)

All data are shown as number (percentage) or mean±SD. HDL indicates high-density lipoprotein; LDL, low-density lipoprotein; N/A, not available; RE-LY, Randomized Evaluation of Long-Term Anticoagulant Therapy; and SPRINT, Systolic Blood Pressure Intervention Trial.

*Other indicates race or ethnicity self-reported as non-White, non-Black and non-Hispanic.

Baseline Risk and Treatment Effect Heterogeneity

In SPRINT, estimated absolute 3.2-year risk for the primary end point with standard treatment (ie, baseline risk) varied widely, from 1% to 50% absolute risk (Figure 1C). There was no significant interaction between baseline risk for the primary end point and the relative treatment effect of intensive treatment in the trial (*P* for interaction=0.92). The baseline risk quartile-specific HRs are shown in Figure 2.

In the RE-LY trial, estimated baseline absolute 1-year risk of stroke or SE, estimated with the GARFIELD-AF risk model, varied from 0.1% to 23.5%, whereas the 1-year risk of major bleeding varied from 0.3% to 13.9% (Figure 1A and 1B, respectively). There was no significant interaction between the baseline risk for the primary efficacy end point and relative treatment effect of either dabigatran, 110 mg, or dabigatran, 150 mg, compared with warfarin (likelihood ratio test *P* for interaction=0.71). There was, however, a significant interaction between baseline risk of major bleeding and the relative treatment effect of both dabigatran, 110 and 150 mg, compared with warfarin (likelihood ratio test *P* for interaction <0.001). The baseline risk quartile-specific

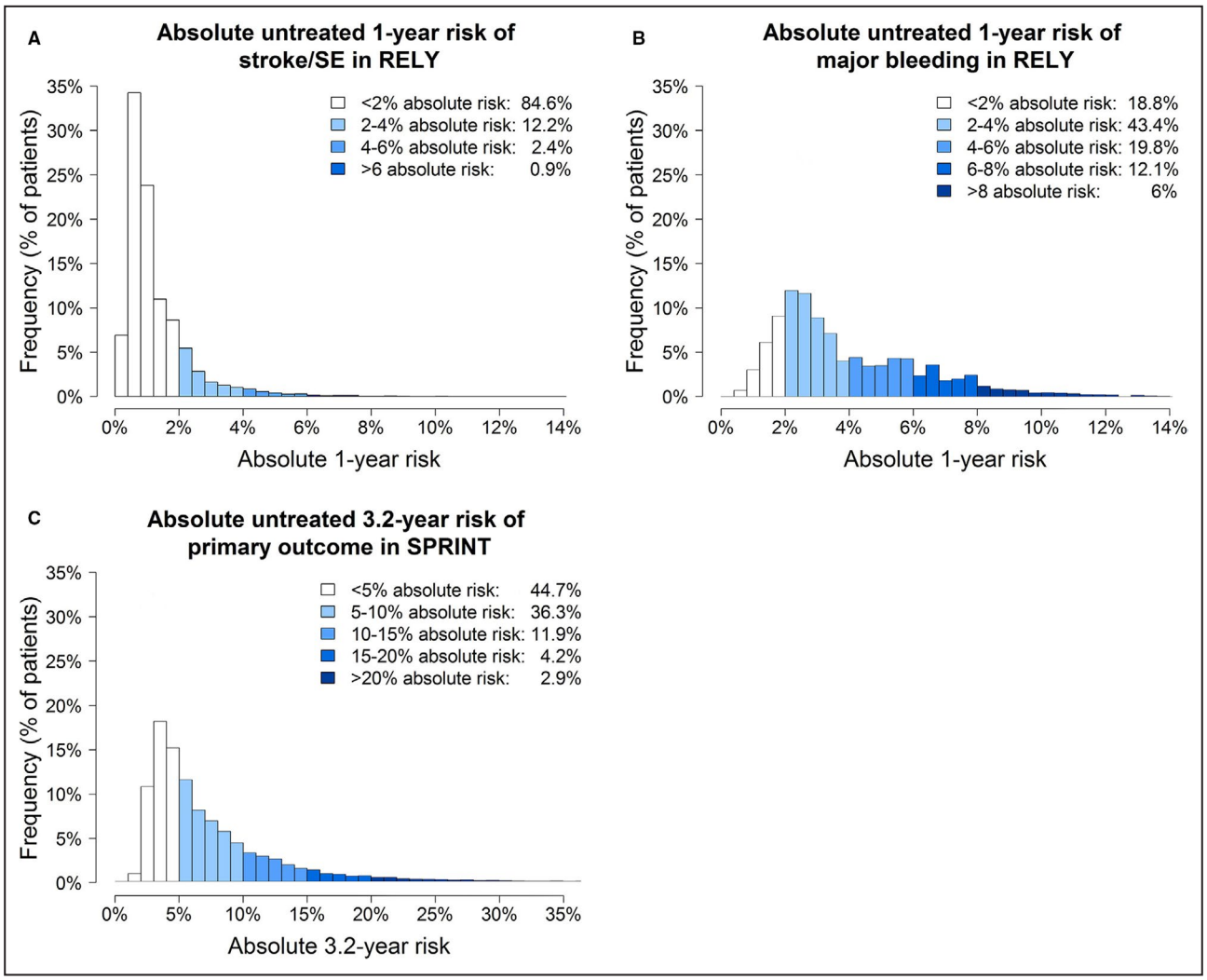

Figure 1. Distribution of untreated (ie, baseline) risk of stroke/systemic embolism (SE) (A) and major bleeding (B) in the RE-LY (Randomized Evaluation of Long-Term Anticoagulant Therapy) trial and of the primary outcome in SPRINT (Systolic Blood Pressure Intervention Trial) (C).

HRs for the primary end point and major bleeding are presented in Figure 3A and 3B, respectively. Quartile-specific HRs ranged from 0.40 (95% CI, 0.26–0.61) to 1.04 (95% CI, 0.83–1.03) for dabigatran, 110 mg, and from 0.61 (95% CI, 0.42–0.88) to 1.20 (95% CI, 0.97–1.50) for dabigatran, 150 mg, compared with warfarin. Baseline characteristics for the quartiles of estimated baseline risk of major bleeding and the quartile-specific HRs for the primary end point according to major bleeding quartiles are shown in Table S1 and Figure S3, respectively. Table S2 shows how randomization remains intact within the risk quartiles.

DISCUSSION

Conventionally, an individual patient's expected absolute treatment effect is calculated from his/her baseline risk and the average relative treatment effect as observed in a clinical trial. However, the relative treatment effect may not be uniform across the trial population. The relative treatment effect may be dependent on an individual's baseline event risk. In the current study, relative treatment effect heterogeneity by an individual's baseline risk was assessed in 2 large randomized clinical trials. In SPRINT, there was no evidence for effect modification of the treatment effect from intensive versus standard blood pressure control on basis of baseline major adverse cardiovascular event risk estimated using a newly derived risk model. In the RE-LY trial, no effect modification for the treatment effect of dabigatran versus warfarin by baseline risk based on the GARFIELD-AF risk model was observed for the risk of the primary end point of stroke and systolic embolism. However, we did establish effect modification by baseline bleeding risk on the treatment effect for the risk of harm from major bleeding, with the lowest baseline bleeding risk quartiles having a clear benefit from dabigatran for both doses, whereas in the highest

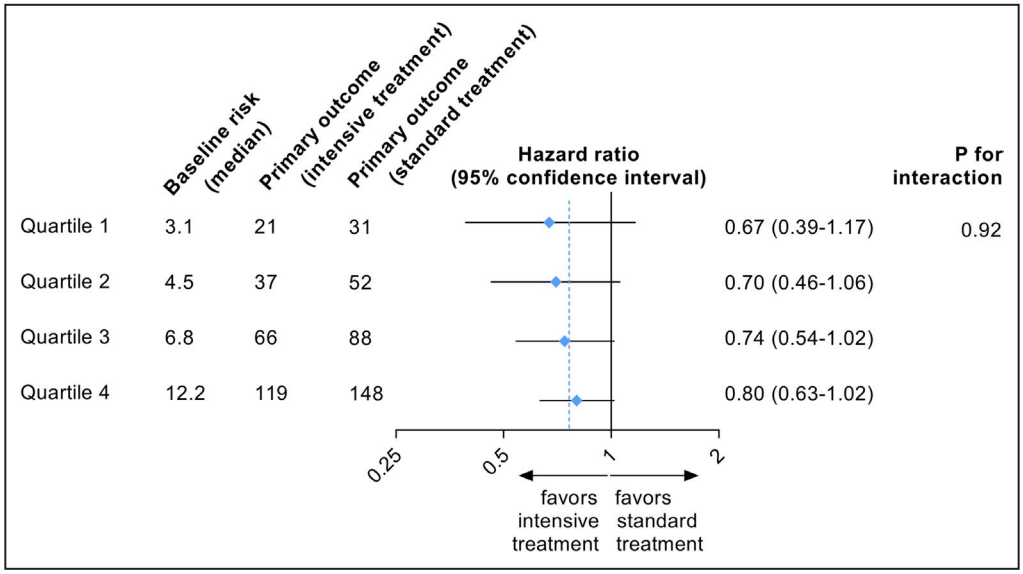

Figure 2. Relative treatment effect in SPRINT (Systolic Blood Pressure Intervention Trial) of intensive versus standard blood pressure control in quartiles of baseline risk for the primary end point.
The blue dotted line denotes the overall trial hazard ratio.

bleeding risk quartiles there is no lower bleeding risk from dabigatran compared with warfarin.

The consequence of this finding is that the average trial result from SPRINT can be applied in all patients. On the basis of the results of the RE-LY trial, the conclusion is that a single relative treatment effect can be used in all patients for the effect of treatment on the primary efficacy end point, but the single relative effect on major bleeding may not be used in all patients. In individualized clinical decision making, from the efficacy standpoint, this may entail offering treatment to all patients, as all will benefit in accordance with their individual baseline risk in combination with the overall HR from the trial (absolute risk reduction). From the safety standpoint, however, when stratifying the study population in quartiles based on their baseline major bleeding risk (ie, the risk when treating with warfarin), there is a clinically important difference in the relative treatment effect in these quartiles. For dabigatran at a dose of 110 mg, a benefit of dabigatran over warfarin is observed in the lowest quartile of predicted bleeding risk, but not in the highest. For dabigatran at a dose of 150 mg, a benefit of dabigatran compared with warfarin, as observed in the lowest 2 quartiles of risk, is offset by a numerical, although just not statistically significant, detrimental effect in the quartile with the highest baseline bleeding risk. This may explain the neutral average main effect on major bleeding reported in the trial (HR, 0.93; 95% CI, 0.81–1.07).[6] The example shows a potential pitfall in the assessment of heterogeneity of treatment effect; looking at efficacy of a treatment alone potentially gives incomplete information necessary for clinical decision making. Assessment of heterogeneity of treatment effects may be most useful if a treatment strategy is costly or confers harms, as physicians may consider withholding such treatment from patients with low expected benefit or important expected harm. Anticoagulation, with a risk of major bleeding, represents an example where this approach is clinically useful. For all therapies, potential treatment benefits should be weighed against potential harms from treatment. For example, a trial investigating strict blood pressure lowering in frail individuals may warrant investigation of treatment heterogeneity on possible adverse effects, such as falls and cognitive decline.

When effect modification is present, further analysis of the data and literature is necessary. In the case of the RE-LY trial, effect modification by both renal function and age has been described in univariate subgroup analyses.[15,16] However, as discussed before, univariate subgroup analyses have disadvantages, including a limited power with a risk of false-positive subgroup finding. Our multivariable approach also identified renal function and age as important factors. A reassuring finding is that interaction analysis between both age and renal function and treatment effect remain statistically significant in an exploratory analysis adjusting for potential confounders (including sex, systolic blood pressure, history of cardiovascular disease, and smoking status; data not shown), making it unlikely that these findings are false positive. The advantage of a multivariable risk-based approach, as described in the current study, includes that the aforementioned “reference class problem” is avoided. For

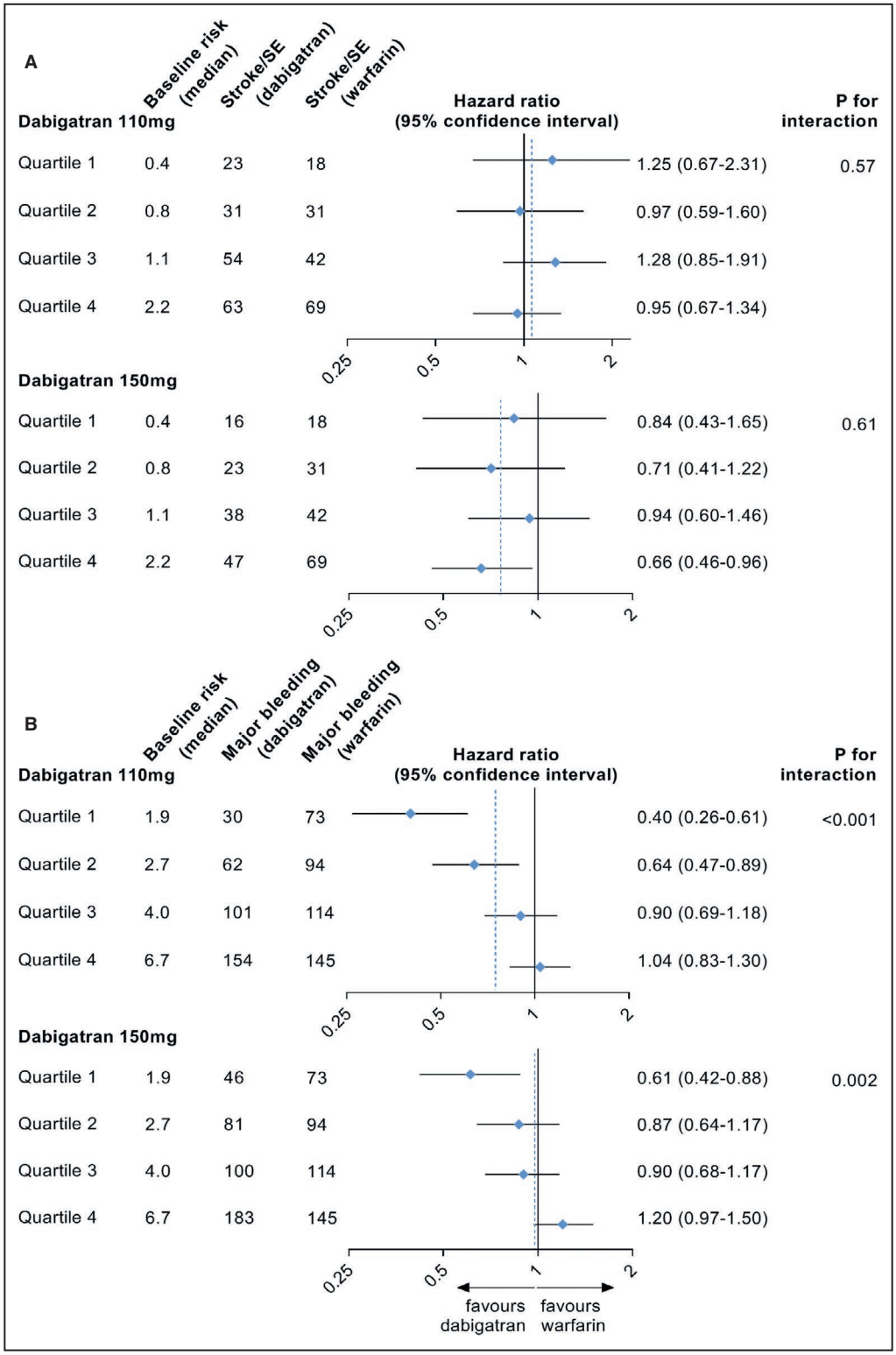

Figure 3. Relative effect in the RE-LY (Randomized Evaluation of Long-Term Anticoagulant Therapy) trial of dabigatran versus warfarin on the risk of stroke/systemic embolism (SE) in quartiles of baseline risk of stroke/SE, according to the Global Anticoagulant Registry in the Field–Atrial Fibrillation (GARFIELD-AF) risk model (A), and major bleeding in quartiles of baseline risk of major bleeding, according to the GARFIELD-AF risk model (B).
The blue dotted line denotes the overall trial hazard ratio.

example, an 80-year-old patient has an increased risk for extracranial major bleeding with either doses of dabigatran compared with warfarin, according to univariate subgroup analyses.[16] At the same time, an individual with a glomerular filtration rate ≥80 mL/min has a remarkably decreased risk of major bleeding, whereas there is a similar risk of bleeding between dabigatran and warfarin in patients with an estimated glomerular filtration rate <50 mL/min.[15] Using just univariate subgroup analyses, it may be difficult to decide whether age or renal function is the more important factor influencing treatment response. A multivariable risk-based approach is therefore more appropriate. This can lead to individual absolute treatment effects in terms of both benefit and harm, which can be weighed in clinical practice to make treatment decisions.

When heterogeneity in treatment effect is found, potentially no single characteristic may be identified that drives treatment effect modification. In that situation, a "one treatment fits all" approach does not apply and simple subgroup analyses do not solve the problem.

A risk model can then be used in clinical practice to determine whether an individual patient qualifies for therapy. For example, patients with atrial fibrillation with a low risk of major bleeding will likely be better off with dabigatran, 150 mg, as they will have a lower risk of major bleeding with dabigatran than with warfarin. At the same time, patients with a higher risk of major bleeding will likely be better off with dabigatran, 110 mg, or even warfarin, as they may have an increased risk of major bleeding with dabigatran, 150 mg, that offsets the reduction in the risk of stroke/SE that is uniform across all patients. The best balance between treatment benefit and treatment disadvantages may be difficult to directly estimate for individual patients based on risk estimation alone. Therefore, decision tools or online calculators should be developed or updated to include personalized treatment effect predictions in cases where risk predictions alone are insufficient for making treatment decisions because of heterogeneity of the treatment effects. These can then be used in daily practice for making informed individualized treatment decisions together with patients.

The advantage of using a preexisting, externally validated risk model, such as GARFIELD-AF, compared with a newly derived risk model, is that it can be used in clinical practice for reliable, unbiased estimates of baseline risk. Furthermore, if relative treatment heterogeneity is present, an existing risk score is likely easier to implement in clinical practice to calculate individual absolute treatment effects for medical decision making. If possible, when designing a new trial, an existing risk model should be selected, so that the appropriate determinants can be collected at baseline. If no risk models exist for the prediction of the primary end point risk, it may be necessary to include the derivation of a new risk model in the trial design if a sufficiently large number of events for model development is anticipated. To deal with the shortcomings of traditional subgroup analyses, we propose that assessment of relative treatment heterogeneity using a multivariable approach should be a prespecified analysis for RCTs. We have shown, using the example from the RE-LY trial, that this method is also applicable in trials with >2 treatment arms (eg, several doses).

There are several additional points to further consider when assessing relative treatment effect heterogeneity. First, when a relative treatment effect is independent of baseline risk, and thus, the same relative effect applies to all patients, the variation in absolute treatment effect may still be large if the range of baseline risk is large.[17] On the other hand, if there is relative treatment heterogeneity, with a significant *P* value for interaction, this does not automatically imply an important treatment heterogeneity. Thus, both the assessment of relative treatment heterogeneity and the estimation of individual absolute treatment effects are of critical importance for the translation of clinical trial results to all patients in clinical practice.

There are limitations to the current method to consider. In the example of SPRINT, an internally derived risk model was used. External validation of this model should be performed before it is applied in clinical practice. However, for the assessment of a risk-treatment interaction, an internally developed model is appropriate, when the internal model performance is sufficient.[18] Thus, the presented model renders unbiased estimates of treatment effect across the spectrum of baseline risk. Furthermore, a possible limitation of the current method is that in trials with strict participant selection criteria, the heterogeneity in patient characteristics may be too small to detect treatment effect modification by baseline risk, and, in addition, the question is whether a risk model derived in a trial population can be generalized to the general population. A careful evaluation of the representativeness of study results is essential for the application of trial results to individuals in clinical practice.

CONCLUSIONS

In conclusion, for translation of an overall trial result into an individual's expected absolute treatment effect, possible differences in relative treatment effect between patients should be taken into account. Effect modification of the relative treatment effect by individual baseline event risk should be assessed systematically in RCTs using multivariable risk prediction, not only in terms of treatment efficacy but also for important treatment harms, as a prespecified analysis. Relative treatment effects can then be translated more reliably to individual absolute treatment effects that can

be weighed for individualized clinical decision making. By using individual patient data in a trial, relative treatment effect modification can be assessed across the full spectrum of risk.

ARTICLE INFORMATION

Received May 14, 2020; accepted June 24, 2021.

Affiliations

Department of Vascular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands (T.I.d.V., M.C. S.-S., J.W., F.L.J.V.); Department of Cardiology, Amsterdam University Medical Center, Academic Medical Center/University of Amsterdam, Amsterdam, the Netherlands (R.J.G.P.); and Julius Center for Health Sciences and Primary Care, Utrecht, the Netherlands (Y.v.d.G.).

Acknowledgments

The authors would like to acknowledge the contributions of Dr Janice Pogue from the Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada, who passed away January 22, 2016. This article was prepared using data from RE-LY (Randomized Evaluation of Long-Term Anticoagulant Therapy) trial data obtained from www.clinicalstudydatarequest.com. The research proposal was reviewed by an Independent Review Panel, and Boehringer Ingelheim did not influence this decision, nor any of the conducted analyses or conclusions. This article was prepared using SPRINT_POP Research Materials obtained from the National Heart, Lung, and Blood Institute (NHLBI) Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the SPRINT_POP or the NHLBI.

Sources of Funding

None.

Disclosures

None.

Supplementary Material

Data S1–S2
Tables S1–S2
Figures S1–S3
References 19,20

REFERENCES

1. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018;363:k4245. doi: 10.1136/bmj.k4245
2. Parker AB, Yusuf S, Naylor CD. The relevance of subgroup-specific treatment effects: the Studies Of Left Ventricular Dysfunction (SOLVD) revisited. *Am Heart J*. 2002;144:941–947. doi: 10.1067/mhj.2002.126446
3. Rothwell PM. Treating individuals 2: subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365:176–186. doi: 10.1016/S0140-6736(05)17709-5
4. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266:93–98. doi: 10.1001/jama.1991.03470010097038
5. Ezekowitz MD, Connolly S, Parekh A, Reilly PA, Varrone J, Wang S, Oldgren J, Themeles E, Wallentin L, Yusuf S. Rationale and design of RE-LY: randomized evaluation of long-term anticoagulant therapy, warfarin, compared with dabigatran. *Am Heart J*. 2009;157:805–810. doi: 10.1016/j.ahj.2009.02.005
6. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, Pogue J, Reilly PA, Themeles E, Varrone J, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2009;361:1139–1151. doi: 10.1056/NEJMoa0905561
7. Ambrosius WT, Sink KM, Foy CG, Berlowitz DR, Cheung AK, Cushman WC, Fine LJ, Goff DC, Johnson KC, Killeen AA, et al. The design and rationale of a multicenter clinical trial comparing two strategies for control of systolic blood pressure: the Systolic Blood Pressure Intervention Trial (SPRINT). *Clin Trials*. 2014;11:532–546. doi: 10.1177/1740774514537404
8. SPRINT Research Group, Wright J, Williamson J, Whelton P, Snyger J, Sink K, Rocco M, Reboussin D, Rahman M, Oparil S, et al. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med*. 2015;373:2103–2116. doi: 10.1056/NEJMoa1511939
9. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837–1847. doi: 10.1161/01.CIR.97.18.1837
10. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetière P, Jousilahti P, Keil U, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J*. 2003;24:987–1003. doi: 10.1016/S0195-668X(03)00114-3
11. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;129:S49–S73. doi: 10.1161/01.cir.0000437741.48606.98
12. Kaasenbrood L, Bhatt DL, Dorresteijn JAN, Wilson PWF, D'Agostino RB, Massaro JM, van der Graaf Y, Cramer MJM, Kappelle LJ, de Borst GJ, et al. Estimated life-expectancy without recurrent cardiovascular events in patients with vascular disease: the SMART-REACH model. *J Am Heart Assoc*. 2018;7:e009217. doi: 10.1161/JAHA.118.009217
13. Jaspers NEM, Ridker PM, Dorresteijn JAN, Visseren FLJ. The prediction of therapy-benefit for individual cardiovascular disease prevention: rationale, implications, and implementation. *Curr Opin Lipidol*. 2018;29:436–444. doi: 10.1097/MOL.0000000000000554
14. Fox KAA, Lucas JE, Pieper KS, Bassand J-P, Camm AJ, Fitzmaurice DA, Goldhaber SZ, Goto S, Haas S, Hacke W, et al. Improved risk stratification of patients with atrial fibrillation: an integrated GARFIELD-AF tool for the prediction of mortality, stroke and bleed in patients with and without anticoagulation. *BMJ Open*. 2017;7:e017157. doi: 10.1136/bmjopen-2017-017157
15. Hijazi Z, Hohnloser SH, Oldgren J, Andersson U, Connolly SJ, Eikelboom JW, Ezekowitz MD, Reilly PA, Siegbahn A, Yusuf S, et al. Efficacy and safety of dabigatran compared with warfarin in relation to baseline renal function in patients with atrial fibrillation: a RE-LY (Randomized Evaluation of Long-term Anticoagulation Therapy) trial analysis. *Circulation*. 2014;129:961–970. doi: 10.1161/CIRCULATIONAHA.113.003628
16. Lauw MN, Eikelboom JW, Coppens M, Wallentin L, Yusuf S, Ezekowitz M, Oldgren J, Nakamya J, Wang J, Connolly SJ. Effects of dabigatran according to age in atrial fibrillation. *Heart*. 2017;103:1015–1023. doi: 10.1136/heartjnl-2016-310358
17. Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int J Epidemiol*. 2016;45:2075–2088. doi: 10.1093/ije/dyw118
18. Burke JF, Hayward RA, Nelson JP, Kent DM. Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. *Circ Cardiovasc Qual Outcomes*. 2014;7:163–169. doi: 10.1161/CIRCOUTCOMES.113.000497
19. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation and Updating*. Springer; 2009.
20. Dalgaard F, Pieper K, Verheugt F, Camm AJ, Fox KAA, Kakkar AK, Pallisgaard JL, Rasmussen PV, Weert HV, Lindhardt TB, et al. GARFIELD-AF model for prediction of stroke and major bleeding in atrial fibrillation: a Danish nationwide validation study. *BMJ Open*. 2019;9:e033283. doi: 10.1136/bmjopen-2019-033283

Supplemental Material

Supplemental Methods

Data S1: *Model derivation and estimation of baseline risk in the SPRINT trial*

A prediction model was derived for the combined outcome (myocardial infarction, other acute coronary syndromes, stroke, heart failure, or death from cardiovascular causes) in 9,361 patients from the SPRINT study population. Prespecified predictors selected on basis of previous risk models and availability in the study data were: age, sex, current smoking, African-American race, history of cardiovascular disease, total cholesterol, HDL cholesterol, estimate glomerular filtration rate (using the CKD-EPI formula), and the use of antihypertensive medication at baseline (9-13). Baseline missing variables were singly imputed using predictive mean matching (aregImpute-algorithm in Rstudio, Hmisc-package). Continuous predictors were truncated at the 1st and 99th percentile to limit the effect of outliers. Whether the association of continuous predictors with the outcome variable is log-linear was assessed with restricted cubic splines; to improve the robustness of the model, transformation was applied when this improved model fit, based on Akaike's Information Criterion (19). Model coefficients of the final model were uniformly shrunken to account for over-optimism with a factor of 4.0% derived from model selection in 1000 bootstrap samples based on the AIC. The model was fitted for the prediction of 3.2-year risk (median follow-up). Model performance was assessed with the c-statistic (95% confidence interval [CI]) for discrimination using 1000 bootstrap samples and with calibration plots of predicted versus observed risk.

This is the underlying formula for the prediction of the risk of the primary outcome:

3.2-year risk: 1 - 0.941 ^ exp(**LP** + 9.416) * 100%

LP = -1.382 * (age in years/10) + 0.137 * ((age in years/10)^2) - 0.234 * (if male) + 0.549 (if current smoker) - 0.092 * (SBP in mmHg) + 0.0003 * (SBP in mmHg)^2 - 0.501 (if history of cardiovascular disease) - 0.700 * (HDL-c in mmol/L) + 0.174 * (total cholesterol in mmol/L) - 0.319 * (eGFR in ml/min) + 0.062 (if African-American) + 0.207 (if currently using antihypertensives) + 0.479 * ((age in years/10) if history of cardiovascular disease) - 0.044 * ((age in years/10)^2 if history of cardiovascular disease)

Data S2: *The GARFIELD-AF risk model*

The Global Anticoagulant Registry in the FIELD-Atrial Fibrillation (GARFIELD-AF) model was developed to allow for simultaneous estimation of all-cause mortality, stroke/systemic embolism (SE), and bleeding risk in patients with newly diagnosed atrial fibrillation (14). This model was found to be superior in predicting stroke/SE and bleeding compared to traditional risk models such as CHA_2DS_2VASc and HAS-BLED (20).

The models were based on the following predictors: age, history of vascular disease, history of stroke, history of bleeding, history of heart failure, chronic kidney disease, region, ethnicity, and oral anticoagulant use.

These are the underlying formulas for the prediction of risk for stroke/SE and bleeding:

1-year risk of ischaemic stroke or systemic embolism = 1 – 0.991344397 ^ exp(**LP**) * 100%

LP = 0.03048226 * (age-60) + 0.952524717 (if history of stroke) + 0.432357326 (if history of bleeding) + 0.319129628 (if heart failure) + 0.574919171 (if chronic kidney disease) + 0.654249546 * (if from Australia, New Zealand or South Africa) + 0.671380382 (if Black / Mixed / Other race (not Caucasian, Hispanic/Latino or Asian)) -0.582045773 (if using Oral Anticoagulant)

1-year risk of haemorrhagic stroke or major bleed = 1 – 0.991344397 ^ exp(**LP**) * 100%

LP = 0.042943 * (age-60) + 0.42205 (if history of vascular disease) + 0.60985 (if chronic kidney disease)

Table S1. Baseline characteristics of the Randomized Evaluation of Long Term Anticoagulant Therapy (RE-LY) trial stratified for quartiles of untreated baseline risk for major bleeding according to the Global Anticoagulant Registry in the FIELD-Atrial Fibrillation (GARFIELD-AF) risk model.

	Quartile 1 n = 5021	**Quartile 2** n = 4204	**Quartile 3** n = 4398	**Quartile 4** n = 4490
Baseline risk of bleeding, median (IQR)	1.9 (1.5-2.0)	2.7 (2.4-2.9)	4.0 (3.6-4.6)	6.7 (5.9-7.9)
Baseline risk of stroke/SE, median (IQR)	0.9 (0.6-1.2)	1.2 (0.8-1.5)	1.6 (1.1-2.4)	2.7 (1.8-3.8)
Age (years)	62 ± 7	72 ± 4	74 ± 5	80 ± 4
Male sex	3533 ± 70	2846 ± 68	2774 ± 63	2361 ± 53
Ethnicity				
Caucasian	3300 (66%)	3171 (75%)	2969 (68%)	3176 (71%)
Black	62 (1%)	33 (1%)	46 (1%)	35 (1%)
Hispanic	225 (5%)	148 (4%)	251 (6%)	255 (6%)
Other	1434 (29%)	852 (20%)	1132 (26%)	1024 (23%)
Current smoking	1750 (35%)	1487 (35%)	1456 (33%)	1286 (29%)
History of cardiovascular disease	217 (4%)	556 (13%)	1232 (28%)	1447 (32%)
Systolic blood pressure (mmHg)	130 ± 17	132 ± 17	131 ± 18	131 ± 18
Body mass index (kg/m^2)	31 ± 6	30 ± 6	28 ± 5	26 ± 5
Estimated glomerular filtration rate ($ml/min/1.73m^2$)	96 ± 29	81 ± 20	65 ± 18	48 ± 10
Use of oral anticoagulants at baseline	3076 (61%)	2773 (66%)	2733 (62%)	2607 (58%)

All data are shown as n (%) or mean ± standard deviation unless stated otherwise; IQR = interquartile range

Table S2. Baseline characteristics of the Randomized Evaluation of Long Term Anticoagulant Therapy (RE-LY) trial stratified for trial allocation *and* quartiles of untreated baseline risk for major bleeding according to the Global Anticoagulant Registry in the FIELD-Atrial Fibrillation (GARFIELD-AF) risk model.

	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
	Standard treatment	**Intensive treatment**	**Standard treatment**	**Intensive treatment**	**Standard treatment**	**Intensive treatment**	**Standard treatment**	**Intensive treatment**
	n = 1181	n = 1160	n = 1155	n = 1185	n = 1164	n = 1176	n = 1183	n = 1157
Age (years)	61 ± 6	61 ± 6	65 ± 7	65 ± 7	69 ± 8	69 ± 8	76 ± 9	76 ± 9
Male sex	554 (47%)	513 (44%)	727 (63%)	723 (61%)	826 (71%)	826 (70%)	928 (78%)	932 (81%)
Ethnicity								
Caucasian	546 (46%)	575 (50%)	633 (55%)	635 (54%)	700 (60%)	690 (59%)	822 (69%)	798 (69%)
Black	472 (40%)	426 (37%)	389 (34%)	387 (33%)	331 (28%)	344 (29%)	231 (20%)	222 (19%)
Hispanic	150 (13%)	144 (12%)	113 (10%)	140 (12%)	113 (10%)	106 (9%)	105 (9%)	113 (10%)
Other	13 (1%)	15 (1%)	20 (2%)	23 (2%)	20 (2%)	36 (3%)	25 (2%)	24 (2%)
Current smoking	53 (4%)	66 (6%)	149 (13%)	173 (15%)	220 (19%)	208 (18%)	180 (15%)	195 (17%)
History of CVD	2 (0%)	3 (0%)	35 (3%)	45 (4%)	210 (18%)	188 (16%)	536 (45%)	543 (47%)
SBP (mmHg)	139 ± 12	139 ± 12	138 ± 14	138 ± 15	140 ± 16	139 ± 16	141 ± 18	142 ± 19
Body mass index (kg/m^2)	31 ± 6	31 ± 6	30 ± 6	30 ± 6	30 ± 5	30 ± 5	28 ± 5	29 ± 5
Total cholesterol (mmol/L)	5.0 ± 0.9	5.0 ± 0.9	5.1 ± 1.0	5.0 ± 1.0	4.9 ± 1.1	4.9 ± 1.1	4.7 ± 1.1	4.8 ± 1.2
LDL cholesterol (mmol/L)	2.9 ± 0.8	2.9 ± 0.8	3.0 ± 0.9	3.0 ± 0.9	2.9 ± 1.0	2.9 ± 1.0	2.8 ± 1.0	2.8 ± 1.0
HDL cholesterol (mmol/L)	1.5 ± 0.4	1.5 ± 0.4	1.4 ± 0.4	1.4 ± 0.3	1.3 ± 0.3	1.3 ± 0.3	1.3 ± 0.3	1.3 ± 0.4
eGFR (ml/min/1.73m^2)	82.1 ± 22.6	82.8 ± 21.9	70.1 ± 21.5	70.9 ± 22.4	62.9 ± 20.2	63.2 ± 20.6	52.5 ± 19.5	52.3 ± 19.6
Statin use at baseline	406 (34%)	372 (32%)	451 (39%)	464 (39%)	559 (48%)	512 (44%)	672 (57%)	647 (56%)
Antihypertensive use	1016 (86%)	978 (84%)	1032 (89%)	1073 (91%)	1065 (91%)	1091 (93%)	1120 (95%)	1104 (95%)

Figure S1. Internal validation of the risk model derived in the Systolic Blood Pressure Intervention Trial (SPRINT) study.

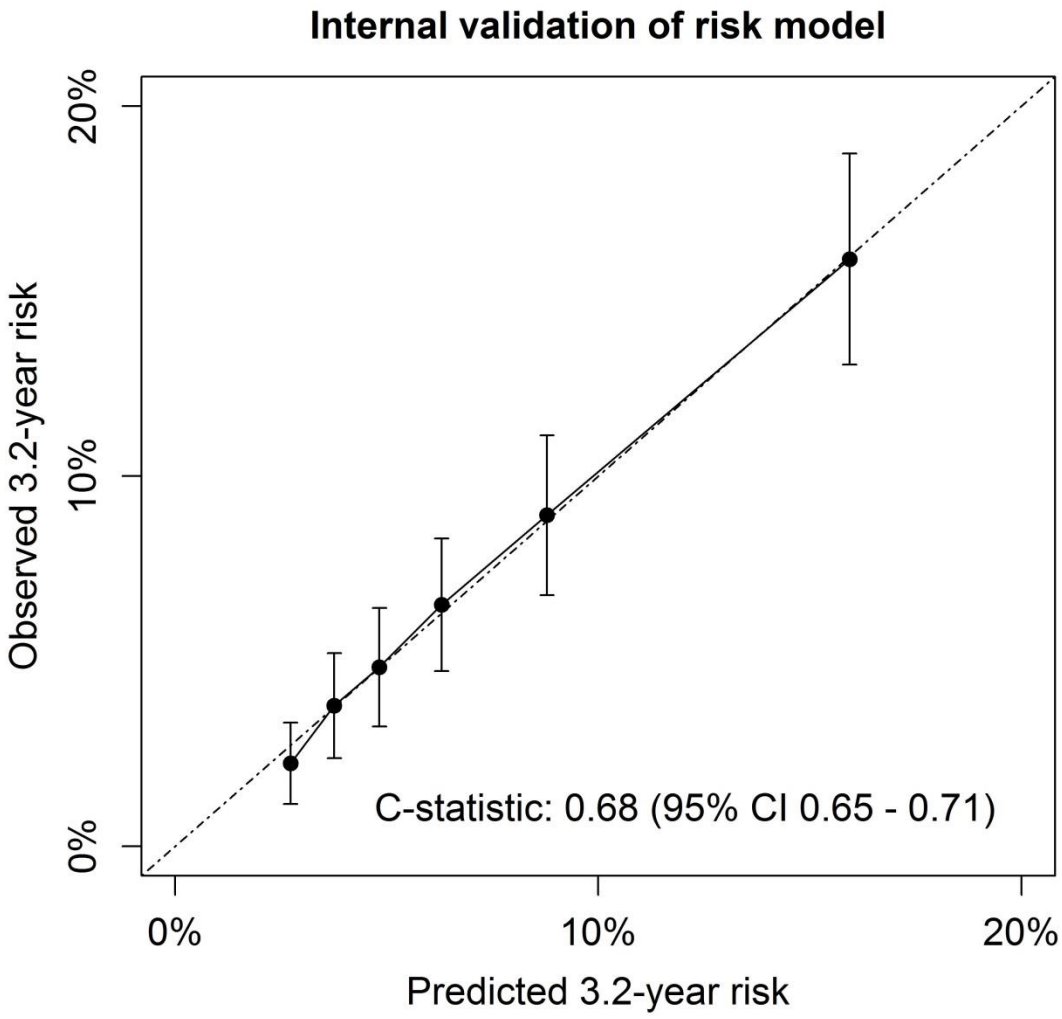

Figure S2. External validation of the Global Anticoagulant Registry in the FIELD-Atrial Fibrillation (GARFIELD-AF) risk model for (A) stroke and systemic embolism, and (b) major bleeding risk in the Randomized Evaluation of Long Term Anticoagulant Therapy (RE-LY) trial.

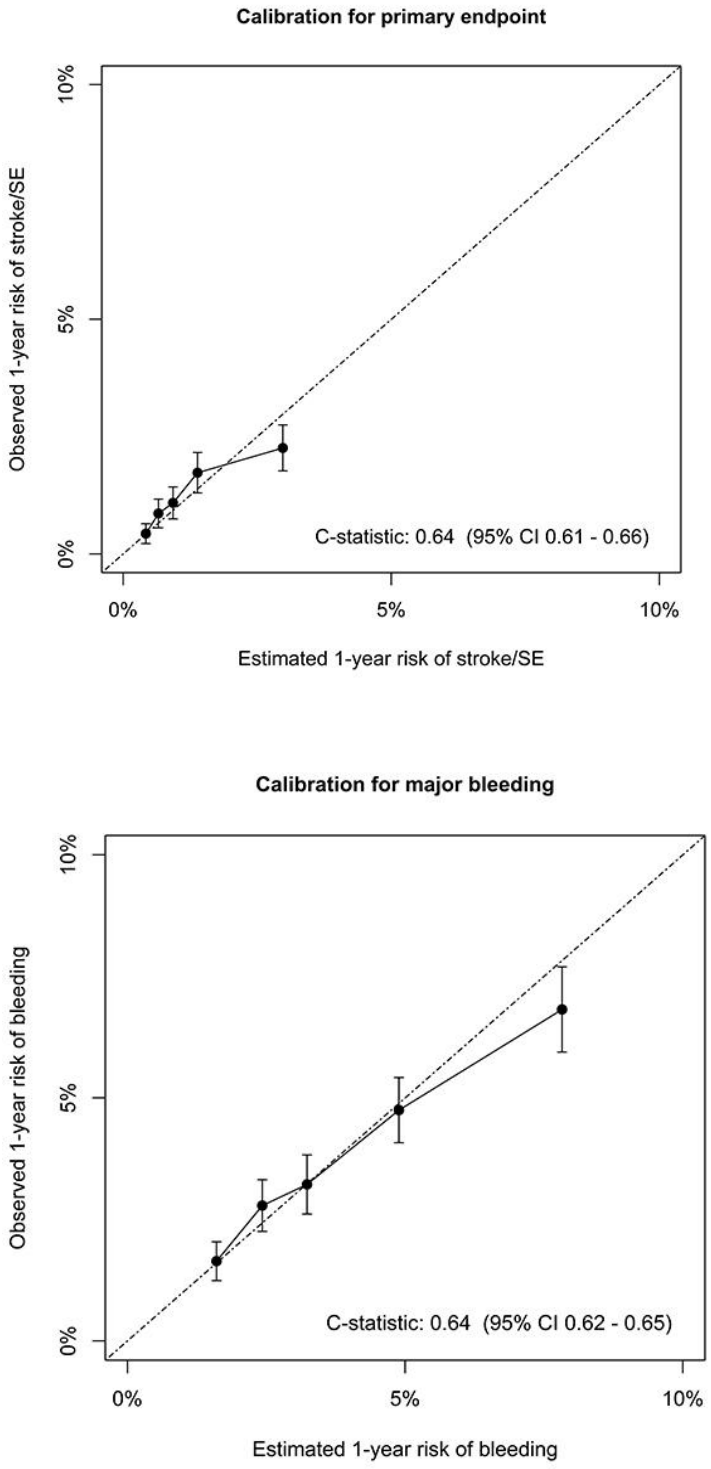

Figure S3. The relative effect of dabigatran (in doses of 110mg or 150mg twice daily) versus warfarin on the risk of major *cardiovascular events* in quartiles of baseline risk of major *bleeding* according to the Global Anticoagulant Registry in the FIELD-Atrial Fibrillation (GARFIELD-AF) risk model.