Accepted Manuscript

Submission Date: 2024-07-17
Accepted Date: 2024-09-09
Accepted Manuscript online: 2024-09-19

Thrombosis and Haemostasis

Risk of recurrent venous thromboembolism in patients with cancer: an individual patient data meta-analysis and development of a prediction model

Vincent Lanting, Toshihiko Takada, Floris Bosch, Andrea Marshall, Michael Grosso, Annie Young, Agnes Y Lee, Marcello Di Nisio, Gary Raskob, Pieter Willem Kamphuisen, HR Buller, Nick van Es.

Affiliations below.

DOI: 10.1055/a-2418-3960

Please cite this article as: Lanting V, Takada T, Bosch F et al. Risk of recurrent venous thromboembolism in patients with cancer: an individual patient data meta-analysis and development of a prediction model. Thromb Haemost 2024. doi: 10.1055/a-2418-3960

Conflict of Interest: M. Grosso is an employee of Daiichi Sankyo. A.Y.Y. Lee reports consulting fees and honoraria from Bayer AG, consulting fees and honoraria from LEO Pharma, consulting fees and honoraria from Pfizer, consulting fees from Servier, and honoraria from Bristol Myers Squibb. M. Di Nisio reports personal fees as an invited speaker from Bayer, Daiichi Sankyo, and Viatris, personal fees for advisory board membership from Leo Pharma, Janssen, and Pfizer, and institutional funding from Leo Pharma, all outside the submitted work. G.E. Raskob reports consultancy fees or honoraria from AMAG Pharma, Alnylam, Anthos Therapeutics, Bayer HealthCare Pharmaceuticals Inc., Bristol-Myers Squibb, Daiichi Sankyo Inc., Ionis, Janssen Global Services LLC, Pfizer, Regeneron, Sirius Pharmaceutical ; honoraria from BMS, Pfizer, Daiichi Sankyo; DSMB or advisory board membership from Anthos Therapetuics, Janssen, Bristol-Myers Squibb and Pfizer, leadership or fiduciary role in other board, society, committee or advocacy group of OU Health, and the National Blood Clot Alliance. P.W. Kamphuisen reports research funding from Daiichi Sankyo and Roche diagnostics. H.R. Büller reports research support from Sanofi-aventis, Bayer HealthCare, Bristol- Myers Squibb, Daiichi-Sankyo, GlaxoSmithKline, Pfizer, Roche, IONIS, Boehringer Ingelheim, Eli Lilly, Novartis. Consultant from Sanofi-aventis, Bayer HealthCare, Bristol- Myers Squibb, Daiichi-Sankyo, GlaxoSmithKline, Pfizer, Roche, IONIS, Boehringer Ingelheim, Eli Lilly, Novartis. Scientific advisory board from Sanofi-aventis, Bayer HealthCare, Bristol-Myers Squibb, Daiichi-Sankyo, GlaxoSmithKline, Pfizer, Roche, IONIS, Boehringer Ingelheim, Eli Lilly, Novartis. N. van Es reports advisory board honoraria from DaiichiSankyo, LEO Pharma, and Bayer, which were transferred to his institute. The other authors have nothing to declare.

Abstract:

Background

About 7% of patients with cancer-associated venous thromboembolism (CAT) develop a recurrence during anticoagulant treatment. Identification of high-risk patients may help guide treatment decisions.

Aim

To identify clinical predictors and develop a prediction model for on-treatment recurrent CAT.

Methods

For this individual patient data (IPD) meta-analysis, we used data from four randomized controlled trials evaluating low-molecular-weight heparin (LMWH) or direct oral anticoagulants (DOACs) for CAT (Hokusai VTE Cancer, SELECT-D, CLOT, and CATCH). The primary outcome was adjudicated on-treatment recurrent CAT during 6-month follow-up. A clinical prediction model was developed using multivariable logistic regression analysis with backward selection. This model was validated using internal-external cross validation. Performance was assessed by the c-statistic and a calibration plot.

Results

After excluding patients using vitamin K antagonists, the combined dataset comprised 2,245 patients with cancer and acute CAT who were treated with edoxaban (23%), rivaroxaban (9%), dalteparin (47%), or tinzaparin (20%). Recurrent on-treatment CAT during 6-month follow-up occurred in 150 (6.7%) patients. Predictors included in the final model were age (restricted cubic spline), breast cancer (OR 0.42; 95%-CI 0.20-0.87), metastatic disease (OR 1.44; 95%-CI 1.01-2.05), treatment with DOAC (OR

0.66; 95%-CI 0.44-0.98), and deep vein thrombosis only as index event (OR 1.72; 95%-CI 1.31-2.27). The c-statistic of the model was 0.63 (95%-CI 0.54-0.72) after internal-external cross validation. Calibration varied across studies.
Conclusions
The prediction model for recurrent CAT included five clinical predictors and has only modest discrimination. Prediction of recurrent CAT at the initiation of anticoagulation remains challenging.

Corresponding Author:
Vincent Lanting, Amsterdam University Medical Centres, Vascular Medicine, Meibergdreef 9, 1105 AZ Amsterdam, Netherlands, v.r.lanting@amsterdamumc.nl

Affiliations:
Vincent Lanting, Amsterdam University Medical Centres, Vascular Medicine, Amsterdam, Netherlands
Vincent Lanting, Tergooi MC, Internal Medicine, Hilversum, Netherlands
Toshihiko Takada, Utrecht University, Utrecht, Netherlands
[...]
Nick van Es, Amsterdam UMC Locatie AMC, Vascular medicine, Amsterdam, Netherlands

Risk of recurrent venous thromboembolism in patients with cancer: an individual patient data meta-analysis and development of a prediction model

Vincent R. Lanting[1,2,3], Toshihiko Takada[4,5], Floris T.M. Bosch[1,2,3], Andrea Marshall[6], Michael A. Grosso[7], Annie M. Young[6,8], Agnes Y.Y. Lee[9], Marcello Di Nisio[10], Gary E. Raskob[11], Pieter W. Kamphuisen[1,2,3], Harry R. Büller[1,2], Nick van Es[1,2]

1. Amsterdam University Medical Center, Department of Vascular Medicine, University of Amsterdam, Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
2. Amsterdam Cardiovascular Sciences, Pulmonary Hypertension & Thrombosis, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
3. Tergooi Hospital, Department of Internal Medicine, Laan van Tergooi 2, 1212 VG, Hilversum, the Netherlands
4. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Universiteitsweg 100, 3584 CG, Utrecht, The Netherlands.
5. Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR), Fukushima Medical University, 1 Hikarigaoka, Fukushima, 960-1247, Japan.
6. University Hospitals Coventry and Warwickshire, Coventry, UK
7. Clinical Development, Daiichi Sankyo, Basking Ridge, New Jersey, USA
8. Warwick Medical School, University of Warwick, Coventry, UK
9. Division of Hematology, University of British Columbia, British Columbia Cancer Agency, Vancouver BC, Canada
10. Department of Medicine and Ageing Sciences, Gabriele D´Annunzio University, Via dei Vestini 31,66100 Chieti, Italy
11. University of Oklahoma Health Sciences Center and OU Health, Oklahoma City, OK, USA

Corresponding author

Vincent Lanting, MD

Amsterdam University Medical Centers, location AMC

Department of vascular medicine, office M0-118

Meibergdreef 9, 1105 AZ, Amsterdam, the Netherlands

Email: v.r.lanting@amsterdamumc.nl

Conflicts interests statement

M. Grosso is an employee of Daiichi Sankyo. **A.Y.Y. Lee** reports consulting fees and honoraria from Bayer AG, consulting fees and honoraria from LEO Pharma, consulting fees and honoraria from Pfizer, consulting fees from Servier, and honoraria from Bristol Myers Squibb. **M. Di Nisio** reports personal fees as an invited speaker from Bayer, Daiichi Sankyo, and Viatris, personal fees for advisory board membership from Leo Pharma, Janssen, and Pfizer, and institutional funding from Leo Pharma, all outside the submitted work. **G.E. Raskob** reports consultancy fees or honoraria from AMAG Pharma, Alnylam, Anthos Therapeutics, Bayer HealthCare Pharmaceuticals Inc., Bristol-Myers Squibb, Daiichi Sankyo Inc., Ionis, Janssen Global Services LLC, Pfizer, Regeneron, Sirius Pharmaceutical ; honoraria from BMS, Pfizer, Daiichi Sankyo; DSMB or advisory board membership from Anthos Therapetuics, Janssen, Bristol-Myers Squibb and Pfizer, leadership or fiduciary role in other board, society, committee or advocacy group of OU Health, and the National Blood Clot Alliance. **P.W. Kamphuisen** reports

research funding from Daiichi Sankyo and Roche diagnostics. **H.R. Büller** reports research support from Sanofi-aventis, Bayer HealthCare, Bristol- Myers Squibb, Daiichi-Sankyo, GlaxoSmithKline, Pfizer, Roche, IONIS, Boehringer Ingelheim, Eli Lilly, Novartis. Consultant from Sanofi-aventis, Bayer HealthCare, Bristol- Myers Squibb, Daiichi-Sankyo, GlaxoSmithKline, Pfizer, Roche, IONIS, Boehringer Ingelheim, Eli Lilly, Novartis. Scientific advisory board from Sanofi-aventis, Bayer HealthCare, Bristol-Myers Squibb, Daiichi-Sankyo, GlaxoSmithKline, Pfizer, Roche, IONIS, Boehringer Ingelheim, Eli Lilly, Novartis. **N. van Es** reports advisory board honoraria from DaiichiSankyo, LEO Pharma, and Bayer, which were transferred to his institute. The other authors have nothing to declare.

Contribution: All authors contributed to the interpretation of the results and writing the manuscript. VL performed data management, analysis, and led writing of the manuscript. TT and NvE performed analysis. FB has written the study protocol and obtained the datasets.

Keywords

- Recurrent venous thromboembolism
- Cancer
- Prediction

Summary Table

What is known on this topic	- Recurrent on-treatment venous thromboembolism is a common complication of cancer-associated thrombosis. - The Ottawa scores are validated scores for prediction of recurrent cancer-associated thrombosis, but the use of the scores in clinical practice is limited due to modest discriminatory ability.
What does this paper add	- This individual patient data meta-analysis of 4 large

	randomized controlled trials identified clinical predictors for recurrent cancer-associated thrombosis before start of anticoagulant treatment. - We derived a clinical prediction model based on age, breast cancer, metastatic disease, treatment with a DOAC, and DVT only as index event. - The model only had modest discriminatory performance, highlighting the need for new risk assessment tools for recurrent cancer-associated thrombosis during treatment.

Abstract

Background

About 7% of patients with cancer-associated venous thromboembolism (CAT) develop a recurrence during anticoagulant treatment. Identification of high-risk patients may help guide treatment decisions.

Aim

To identify clinical predictors and develop a prediction model for on-treatment recurrent CAT.

Methods

For this individual patient data (IPD) meta-analysis, we used data from four randomized controlled trials evaluating low-molecular-weight heparin (LMWH) or direct oral anticoagulants (DOACs) for CAT (Hokusai VTE Cancer, SELECT-D, CLOT, and CATCH). The primary outcome was adjudicated on-treatment recurrent CAT during 6-month follow-up. A

clinical prediction model was developed using multivariable logistic regression analysis with backward selection. This model was validated using internal-external cross validation. Performance was assessed by the c-statistic and a calibration plot.

Results

After excluding patients using vitamin K antagonists, the combined dataset comprised 2,245 patients with cancer and acute CAT who were treated with edoxaban (23%), rivaroxaban (9%), dalteparin (47%), or tinzaparin (20%). Recurrent on-treatment CAT during 6-month follow-up occurred in 150 (6.7%) patients. Predictors included in the final model were age (restricted cubic spline), breast cancer (OR 0.42; 95%-CI 0.20-0.87), metastatic disease (OR 1.44; 95%-CI 1.01-2.05), treatment with DOAC (OR 0.66; 95%-CI 0.44-0.98), and deep vein thrombosis only as index event (OR 1.72; 95%-CI 1.31-2.27). The c-statistic of the model was 0.63 (95%-CI 0.54-0.72) after internal-external cross validation. Calibration varied across studies.

Conclusions

The prediction model for recurrent CAT included five clinical predictors and has only modest discrimination. Prediction of recurrent CAT at the initiation of anticoagulation remains challenging.

Introduction

Venous thromboembolism (VTE), comprising deep-vein thrombosis (DVT) and pulmonary embolism (PE), is a frequent complication in patients with cancer.[1] Direct oral anticoagulants (DOACs) or low-molecular-weight heparin (LMWH) are recommended for the treatment of acute VTE,[2-6] but the risk of recurrence nonetheless remains high.[7] In a meta-analysis of six randomized controlled trials (RCTs), the cumulative incidences of recurrent VTE over a 6-

months treatment period were 5.4% and 8.3% in patients receiving DOAC or LMWH, respectively.[8]

Patients with cancer and acute VTE are usually treated for at least 3-6 months. Anticoagulation is usually continued in case of active cancer or ongoing anticancer treatment. Decisions about the optimal intensity and duration of anticoagulant treatment should ideally be guided by the risk recurrent VTE. For example, while in the RCTs the dose of LMWH was typically reduced by 25% after the first month of treatment to mitigate the risk of bleeding, but it is unknown if this dose reduction strategy should be avoided in cancer patients at high risk of recurrent VTE. Currently, the only risk stratification tool to determine the risk of recurrent VTE in cancer patients is the Ottawa score, which stratifies the risk of recurrence based on tumor type, cancer stage, and history of VTE.[9] However, several studies have shown poor discrimination of this score (c-statistics ranging from 0.5 to 0.7), which has limited its use in clinical practice.[10,11] In addition, this score provides a risk classification rather than an individualized risk estimate. Therefore, we sought to derivate and validate a novel clinical prediction model for recurrent VTE in cancer patients with acute VTE.

Methods

Study selection

This report adheres to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidance for IPD-meta-analysis (**Supplementary Table 1)**.[12] We identified RCTs that evaluated anticoagulant treatment in patients with cancer and acute VTE up to 2021 based on previously published systematic reviews [7,13]. Studies were eligible if they included adult patients with active cancer (other than basal-cell

or squamous-cell skin cancer) and acute symptomatic or incidental DVT or PE, and had at least 6 months of follow-up. Of eight identified trials[2-6,14-16] (**Supplementary Table 2**), individual patient data (IPD) were obtained from four studies: Hokusai VTE cancer trial [2], SELECT-D[3], CATCH[14], and CLOT[15]. These trials enrolled patients between 1999 and 2016. In all studies, active cancer was defined as a cancer diagnosis or cancer treatment in the 6 months prior to the first VTE event, or the presence of recurrent, regionally advanced, or metastatic solid cancer, or hematological cancer not in remission. The primary efficacy outcome was symptomatic or incidentally detected recurrent VTE in Hokusai VTE cancer and SELECT-D, while only symptomatic events were considered in the primary efficacy outcome of the CLOT and CATCH studies. In CLOT and CATCH, a vitamin K antagonist was compared with LMWH (dalteparin and tinzaparin, respectively), while Hokusai VTE cancer and SELECT-D trials compared an oral factor Xa inhibitor (edoxaban or rivaroxaban, respectively) with LMWH (dalteparin). Since vitamin K antagonists are no longer recommended as treatment for cancer-associated thrombosis,[17-19] patients allocated to these agents were excluded from the present analysis. The primary outcome was recurrent on-treatment VTE, which was defined a symptomatic or incidentally detected DVT or PE that was diagnosed during use of study treatment. In the original studies, all outcome events were adjudicated without knowledge of treatment allocation. In the present analysis, only events that were adjudicated by the original study as being on-treatment were included. The definition of the on-treatment period was from randomization until 24-72 hours after last intake of study drug.

Selection of candidate predictors and model development

Candidate predictors were selected based on their known association with a first or recurrent VTE in the literature and their availability in the databases.[20-22] Based on the (modified) Ottawa score, breast and lung cancer were evaluated as binary predictors. In

addition, we also evaluated cancer types associated with the risk of a first VTE, including hepatobiliary cancer, gynecological cancer, hematological cancer, and genitourinary cancer excluding prostate cancer. In an explorative analysis, cancer type was categorized based on the risk of a first VTE using the classification proposed by Li and colleagues which includes[23] very high-risk cancer (pancreatic, gastroesophageal, bile duct, and gall bladder cancer), high-risk cancer (lung, ovarian, uterine, bladder, kidney, testicular, primary brain cancer, aggressive non-Hodgkin lymphoma, multiple myeloma, and soft tissue sarcoma), intermediate-risk cancer (colorectal cancer), and low-risk cancer (all other cancers). Other candidate predictors included age (continuous), sex, body weight (continuous), platelet count of >350 $x10^9$/L, use of antiplatelet agents, type of anticoagulant treatment (LMWH vs DOAC), and index VTE type (PE with or without DVT vs DVT only). The following candidate predictors were identified but could not be used because they were not available in all databases: hemoglobin level, leukocyte count, smoking, ethnicity, anti-cancer treatment, and plasma creatinine. Partially missing data for candidate predictors up to 15% were imputed within studies using multiple imputation with chained equations, using a model that included most baseline variables as well as outcomes.[24] Systematically missing data were not imputed.

Candidate predictors were first evaluated in a univariable logistic regression model within each study. Odds ratios were pooled in a random effects meta-analysis using the Hartung-Knapp method. Between-study heterogeneity was assessed for each predictor and displayed using forest plots. Variables were used for model development if there was no evidence of substantial heterogeneity. These candidate predictors were subsequently included in a multivariable logistic regression ('full model'). Restricted cubic splines restricted to 3 knots were used to evaluate whether transformation of continuous variables

was appropriate. Variables in the final model were selected using stepwise backward selection using Akaike's information criterion (AIC; P<0.157).[25] Discrimination of the model was evaluated by calculating the c-statistic. The c-statistic can be calculated by using all possible pairs of patients where one patient experienced VTE and the other patient did not. The c-statistic is the proportion of such pairs in which the patient with VTE had a higher predicted probability of experiencing VTE than the subject who did not have VTE. Calibration was assessed by calculating the ratio between the number of observed and expected events (O:E ratio) and a calibration plot in each study. Ideally, the O:E ratio should be 1. If the OE ratio is <1, the model overestimates the probability of having recurrent VTE. If the O:E ratio is >1, the model underestimates the probability of having recurrent VTE. The model was validated using internal-external cross-validation, in which a new model was iteratively derived in n-1 studies and subsequently evaluated in the remaining study. Performance measures were pooled across the internal-external cross validation iterations by a random-effects meta-analysis with restricted maximum likelihood estimation and the Hartung-Knapp-Sidik-Jonkman method to calculate 95% confidence intervals (CI).[26] Prediction intervals were calculated as a measure of between-study heterogeneity, which indicates expected model performance when the prediction model is applied within a specific study. All analyses were performed using R, version 2.2.1 (www.R-project.org).

Results

Characteristics of study group

Data from Hokusai VTE Cancer (n=1,046), SELECT-D (n=406), CLOT (n=676), and CATCH (n=914) were used (see **Supplementary Table 2** for study details). These trials enrolled patients from North-America, Europe, and Oceania. After exclusion of patients treated with

vitamin K antagonists from CLOT and CATCH, the combined IPD set comprised 2,245 patients. The mean age was 63 years (standard deviation [SD], 12) and 51% were female (**Table 1**). The most frequent cancer types were colorectal (17%), lung (13%), and breast cancer (12%) (**Supplementary Table 3**). At randomization, 1,300 patients (59%) had metastatic cancer. Patients were randomly allocated to edoxaban (23%), rivaroxaban (9%), dalteparin (47%), or tinzaparin (20%). During 6 months of follow-up, 150 (6.7%) patients developed on-treatment recurrent VTE including PE with or without DVT (54%), DVT only (45%), or other VTE (1%), and 30.4% died (**Table 1**).

Candidate predictors

Supplementary Figure 2 and supplementary table 5 show the association between the 15 candidate predictors and recurrent VTE in each study. **Table 2** shows the results from the univariable logistic regression model. The candidate predictors with the strongest association with recurrent VTE were DVT only at randomization (OR 1.80; 95% CI: 1.29-2.52, I^2=0%), breast cancer (OR 0.41; 95% CI: 0.20-0.84, I^2=0%), and treatment with a DOAC (OR 0.57; 95% CI: 0.38-0.85, I^2=0%) (**Table 2**).

Prediction model

All candidate predictors were included in the full model. After stepwise backward selection, the following five predictors were retained in the final multivariable logistic regression model: age (continuous), breast cancer, metastatic disease, DOAC or LMWH treatment, and DVT only as index event (**Table 2;** formula provided in **Supplementary Table 4**). The pooled apparent c-statistic of the model was 0.66 (95% CI: 0.61-0.70), which decreased to 0.63 (95% CI: 0.54-0.72; 95%, prediction interval: 0.22-0.91) after internal-external cross validation

(**Figure 1**). Calibration-in-the-large was good with a ratio between observed and expected outcomes of 1.01 (95%-CI: 0.85-1.21) (**Figure 2**). Calibration across the studies varied though (**Supplementary Figure 1**), with poor calibration in the CLOT and CATCH trials and better calibration in the Hokusai VTE Cancer and SELECT-D. Specifically, the model underestimated recurrent VTE risk in SELECT-D trial and overestimated the risk in the CATCH trial.

Discussion

Using IPD from four RCTs including more than 2,000 patients with cancer and acute VTE, five clinical predictors of recurrent on-treatment VTE were identified. The strongest predictors were DVT only (OR, 1.80), breast cancer (OR, 0.41), and treatment with a DOAC compared to LMWH (OR, 0.57). The clinical prediction model for the 6-month risk of on-treatment recurrent VTE including these five predictors had modest discrimination (c-statistic 0.63 after internal-external cross validation) and calibration was inconsistent.

The Ottawa risk score is currently the only validated tool for assessment of the risk of recurrence after cancer-associated VTE.[11] The score's items include sex, previous VTE, cancer stage, and cancer type (breast or lung cancer). Two versions of the score have been developed: the original score that classifies patients as low or high risk, while the modified Ottawa score also includes an intermediate risk group. Unfortunately, we were not able to formally evaluate the performance of the Ottawa scores since data on TNM classification were not collected in all RCTs. A systematic review and meta-analysis demonstrated that discrimination of the original (c-statistic 0.7; 95% CI: 0.6-0.8) and modified Ottawa scores (c-statistic 0.5; 95% CI: 0.5-0.5) is comparable to that of the clinical prediction model presented here.[11]

Another prediction model for cancer-associated recurrent VTE was recently developed using Spanish electronic health record data from 16,407 cancer patients.[27] After feature selection and model training using machine learning, the items included in the model were age, previous VTE, VTE type, metastasis, adenocarcinoma, hemoglobin and serum creatinine levels, and platelet and leukocyte count. Discrimination of the model was also modest, with c-statistics ranging between 0.66 and 0.69 depending on the statistical technique used. Although this retrospective derivation study was well-powered, it is unclear how many events occurred during anticoagulant treatment and what the positive predictive value of the administrative codes used for recurrent VTE was. The model has not been externally validated yet. Unfortunately, we were also unable to validate this model due to missing information in our dataset, in particular several laboratory data were not available.

Tumor type is by far the strongest predictor for a first episode of cancer-associated VTE, but the prognostic value of tumor type for recurrent VTE is less clear.[28] A large Danish population-based cohort including 34,072 patients with cancer and a first VTE diagnosis, identified cancer type as a predictor for recurrent VTE, but the associations were generally weak.[29] The strongest association were observed for genitourinary (subdistribution hazard ratio [HR] 1.35; 95% CI:1.06- 1.71) and lung cancer (subdistribution HR 1.26; 95% CI:1.03-1.53). In the present study, only breast cancer was retained as a protective risk factor in the final model for recurrent VTE. Discrimination was not improved when the validated tumor risk classification for a first VTE proposed by Li and colleagues was used.[23] Similarly, cancer type was not retained in the aforementioned model by Munoz and colleagues. These findings suggests that the association between cancer type and a first VTE is stronger than

that with a recurrent VTE, a similar phenomenon previously observed for hereditary thrombophilia that has been attributed to collider bias.[30] Whether a specific cancer type risk classification for recurrent VTE improves prediction needs further study.

The current study had several strengths. We were able to obtain high quality patient-level data from the four open-label RCTs that were reasonably homogeneous in design and outcome definitions. The proportion of missing data was low, few patients were lost to follow-up, and all recurrent thromboembolic events were adjudicated. The number of outcome events per variable included in the full model was about 27, which is generally believed to be sufficient for model development. The internal-external cross-validation procedure allowed us to validate the model using all available data unlike a split-sample approach.

Some limitations merit consideration. First, we were not able to assess other potential predictors of recurrent VTE, such as cancer stage, kidney function, hemoglobin levels, leukocyte count, history of VTE, and cancer treatment, as they were missing in one or more studies. Platelet count had to be used dichotomously because continuous data were not available in all studies. Second, we could not directly compare the performance of the present model to other previously developed risk assessment tools such as the Ottawa score, because of missing predictors in our database. Third, we did not have access to data from more recent trials, such as CANVAS or Caravaggio.[5,6] Fourth, we only used data from RCTs which can limit generalizability. The strict eligibility criteria used in the clinical trials likely resulted in patients with a better prognosis than in the general population, with unclear potential effect on the performance of the model. External validation of the model in

other settings would be needed. Fifth, participants in CATCH and CLOT were enrolled more than 10 to 20 years ago respectively, with resulting differences in cancer treatment, follow-up (e.g. staging scans), and diagnostic procedures for VTE compared with the Hokusai VTE cancer and SELECT-D trials. Also, there was some variation in the definition of recurrent VTE across the trials. In CLOT, incidental VTE was not considered in the primary outcome. Hokusai VTE cancer and CATCH adjudicated unexplained death as fatal PE, since PE could not be ruled out. These differences may have led to the poor calibration observed in the CATCH and CLOT trials. Furthermore, the discriminatory ability of the final model was lower in the CATCH trial compared with the other three trials, which might be explained by differences in case mix (e.g. differences in cancer type with other recurrent VTE rates), differences in treatment (e.g. full-dose LMWH in CATCH control group compared to maintenance dose LMWH in the other trials), differences in outcome definition (about half of recurrent VTE in CATCH were deaths for which PE could not be ruled out), or just chance.

Discrimination of the present prediction model for recurrent VTE was not better than that of the (modified) Ottawa score nor the model by Muñoz et al.[11,27] Discrimination of all these models is modest at best (c-statistics ≤0.70), but comparable to performance of a prediction model for recurrent VTE in the general population.[31] Prediction of recurrent VTE is challenging because it is often provoked by factors that occur during anticoagulant treatment, such as surgery, changes in systemic anticancer therapy, hospitalization for an acute medical illness, or cancer progression. Other contributing factors include interruptions of anticoagulation for surgery or bleeding and adherence, which may be lower for LMWH than for DOACs. Such factors cannot be incorporated in statistical prediction models that are applied only once at baseline. Dynamic prediction models can overcome this limitation by

allowing periodic reassessment, but they are much harder to develop and validate. Extending the clinical model with plasma biomarkers, such as soluble P-selectin, may improve prediction at start of anticoagulation at the cost of adding complexity.[32]

Another important point is the timing of applying a prediction model to guide treatment decisions. Patients classified as being at high risk of recurrent at the index VTE should probably not have a LMWH dose reduction at 1 month, but it is less clear if such patients should also continue full-dose anticoagulation beyond 3-6 months. Ideally, a new assessment at 3-6 months is needed to guide this decision, which is of particular interest given the upcoming studies that evaluate a low-dose DOAC for secondary prevention in cancer patients, such as the API-CAT trial (NCT03692065) and EVE trial, as well as trials evaluating factor XI inhibitors.[33] Accurate prediction of recurrent VTE at different time points during the course of the disease remains an important unmet need.

In conclusion, we have developed a prediction model with five predictors using the IPD of four randomized controlled trials. However, discrimination of the final clinical prediction model was modest, indicating that prediction of cancer-associated recurrent VTE at diagnosis of acute VTE remains challenging and that other contributing factors need to be identified.

Acknowledgments

This study is based on research using data from data contributors Pfizer that has been made available through Vivli, Inc. Vivli has not contributed to or approved, and is not in any way responsible for, the contents of this publication. The SELECT-D trial was supported by an

unrestricted educational grant from Bayer AG, which also provided rivaroxaban and placebo tablets. The CATCH trial was sponsored and funded by LEO Pharma. The Hokusai VTE cancer trial was supported by Daiichi Sankyo.

References

1. Timp JF, Braekkan SK, Versteeg HH, Cannegieter SC. Epidemiology of cancer-associated venous thrombosis. *Blood*. Sep 5 2013;122(10):1712-23. doi:10.1182/blood-2013-04-460121
2. Raskob GE, van Es N, Verhamme P, et al. Edoxaban for the Treatment of Cancer-Associated Venous Thromboembolism. *N Engl J Med*. Feb 15 2018;378(7):615-624. doi:10.1056/NEJMoa1711948
3. Young AM, Marshall A, Thirlwall J, et al. Comparison of an Oral Factor Xa Inhibitor With Low Molecular Weight Heparin in Patients With Cancer With Venous Thromboembolism: Results of a Randomized Trial (SELECT-D). *J Clin Oncol*. Jul 10 2018;36(20):2017-2023. doi:10.1200/JCO.2018.78.8034
4. McBane RD, 2nd, Wysokinski WE, Le-Rademacher JG, et al. Apixaban and dalteparin in active malignancy-associated venous thromboembolism: The ADAM VTE trial. *J Thromb Haemost*. Feb 2020;18(2):411-421. doi:10.1111/jth.14662
5. Agnelli G, Becattini C, Meyer G, et al. Apixaban for the Treatment of Venous Thromboembolism Associated with Cancer. *New England Journal of Medicine*. 2020;382(17):1599-1607. doi:10.1056/NEJMoa1915103
6. Schrag D, Uno H, Rosovsky R, et al. Direct Oral Anticoagulants vs Low-Molecular-Weight Heparin and Recurrent VTE in Patients With Cancer: A Randomized Clinical Trial. *JAMA*. 2023;329(22):1924-1933. doi:10.1001/jama.2023.7843
7. Mulder FI, Bosch FTM, Young AM, et al. Direct oral anticoagulants for cancer-associated venous thromboembolism: a systematic review and meta-analysis. *Blood*. 2020/09/17/ 2020;136(12):1433-1441. doi:https://doi.org/10.1182/blood.2020005819
8. Frere C, Farge D, Schrag D, Prata PH, Connors JM. Direct oral anticoagulant versus low molecular weight heparin for the treatment of cancer-associated venous thromboembolism: 2022 updated systematic review and meta-analysis of randomized controlled trials. *J Hematol Oncol*. May 21 2022;15(1):69. doi:10.1186/s13045-022-01289-1
9. Louzada ML, Carrier M, Lazo-Langner A, et al. Development of a clinical prediction rule for risk stratification of recurrent venous thromboembolism in patients with cancer-associated venous thromboembolism. *Circulation*. Jul 24 2012;126(4):448-54. doi:10.1161/circulationaha.111.051920
10. Mulder FI, Kraaijpoel N, Di Nisio M, et al. The Ottawa score performs poorly in cancer patients with incidental pulmonary embolism. *Thromb Res*. Sep 2019;181:59-63. doi:10.1016/j.thromres.2019.07.005
11. Delluc A, Miranda S, Exter PD, et al. Accuracy of the Ottawa score in risk stratification of recurrent venous thromboembolism in patients with cancer-associated venous thromboembolism: a systematic review and meta-analysis. *Haematologica*. May 2020;105(5):1436-1442. doi:10.3324/haematol.2019.222828

12. Debray TPA, Collins GS, Riley RD, Snell KIE, Van Calster B, Reitsma JB, Moons KGM. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. *Bmj*. Feb 7 2023;380:e071018. doi:10.1136/bmj-2022-071018
13. Rossel A, Robert-Ebadi H, Combescure C, et al. Anticoagulant therapy for acute venous thrombo-embolism in cancer patients: A systematic review and network meta-analysis. *PLoS One*. 2019;14(3):e0213940. doi:10.1371/journal.pone.0213940
14. Lee AYY, Kamphuisen PW, Meyer G, et al. Tinzaparin vs Warfarin for Treatment of Acute Venous Thromboembolism in Patients With Active Cancer: A Randomized Clinical Trial. *JAMA*. 2015;314(7):677-686. doi:10.1001/jama.2015.9243
15. Lee AYY. Low-Molecular-Weight Heparin versus a Coumarin for the Prevention of Recurrent Venous Thromboembolism in Patients with Cancer. *N Engl J Med*. 2003:349:146-53.
16. Planquette B, Bertoletti L, Charles-Nelson A, et al. Rivaroxaban vs Dalteparin in Cancer-Associated Thromboembolism: A Randomized Trial. *Chest*. Mar 2022;161(3):781-790. doi:10.1016/j.chest.2021.09.037
17. Farge D, Frere C, Connors JM, et al. 2022 international clinical practice guidelines for the treatment and prophylaxis of venous thromboembolism in patients with cancer, including patients with COVID-19. *Lancet Oncol*. Jul 2022;23(7):e334-e347. doi:10.1016/s1470-2045(22)00160-7
18. Lyman GH, Carrier M, Ay C, et al. American Society of Hematology 2021 guidelines for management of venous thromboembolism: prevention and treatment in patients with cancer. *Blood Adv*. Feb 23 2021;5(4):927-974. doi:10.1182/bloodadvances.2020003442
19. Falanga A, Ay C, Di Nisio M, et al. Venous thromboembolism in cancer patients: ESMO Clinical Practice Guideline☆. *Annals of Oncology*. 2023/05/01/ 2023;34(5):452-467. doi:https://doi.org/10.1016/j.annonc.2022.12.014
20. Chee CE, Ashrani AA, Marks RS, Petterson TM, Bailey KR, Melton LJ, 3rd, Heit JA. Predictors of venous thromboembolism recurrence and bleeding among active cancer patients: a population-based cohort study. *Blood*. Jun 19 2014;123(25):3972-8. doi:10.1182/blood-2014-01-549733
21. Cohen AT, Katholing A, Rietbrock S, Bamber L, Martinez C. Epidemiology of first and recurrent venous thromboembolism in patients with active cancer. A population-based cohort study. *Thromb Haemost*. Jan 5 2017;117(1):57-65. doi:10.1160/th15-08-0686
22. Hansson PO, Sörbo J, Eriksson H. Recurrent venous thromboembolism after deep vein thrombosis: incidence and risk factors. *Arch Intern Med*. Mar 27 2000;160(6):769-74. doi:10.1001/archinte.160.6.769
23. Li A, La J, May SB, et al. Derivation and Validation of a Clinical Risk Assessment Model for Cancer-Associated Thrombosis in Two Unique US Health Care Systems. *J Clin Oncol*. Jan 10 2023:Jco2201542. doi:10.1200/jco.22.01542
24. Burgess S, White IR, Resche-Rigon M, Wood AM. Combining multiple imputation and meta-analysis with individual participant data. *Stat Med*. Nov 20 2013;32(26):4499-514. doi:10.1002/sim.5844
25. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol*. Oct 2004;53(5):793-808. doi:10.1080/10635150490522304

26. van Es N, Takada T, Kraaijpoel N, et al. Diagnostic management of acute pulmonary embolism: a prediction model based on a patient data meta-analysis. *Eur Heart J*. Aug 22 2023;44(32):3073-3081. doi:10.1093/eurheartj/ehad417
27. Muñoz AJ, Souto JC, Lecumberri R, et al. Development of a predictive model of venous thromboembolism recurrence in anticoagulated cancer patients using machine learning. *Thrombosis Research*. 2023/08/01/ 2023;228:181-188. doi:https://doi.org/10.1016/j.thromres.2023.06.015
28. Mulder FI, Horváth-Puhó E, van Es N, et al. Venous thromboembolism in cancer patients: a population-based cohort study. *Blood*. Apr 8 2021;137(14):1959-1969. doi:10.1182/blood.2020007338
29. Ording AG, Nielsen PB, Skjøth F, et al. Risk of recurrent cancer-associated venous thromboembolism: A Danish nationwide cohort study. *International Journal of Cardiology*. 2023/11/01/ 2023;390:131271. doi:https://doi.org/10.1016/j.ijcard.2023.131271
30. Digitale JC, Martin JN, Glidden DV, Glymour MM. Key concepts in clinical epidemiology: collider-conditioning bias. *Journal of Clinical Epidemiology*. 2023;161:152-156. doi:10.1016/j.jclinepi.2023.07.004
31. de Winter MA, Büller HR, Carrier M, et al. Recurrent venous thromboembolism and bleeding with extended anticoagulation: the VTE-PREDICT risk score. *Eur Heart J*. Apr 7 2023;44(14):1231-1244. doi:10.1093/eurheartj/ehac776
32. van Es N, Louzada M, Carrier M, et al. Predicting the risk of recurrent venous thromboembolism in patients with cancer: A prospective cohort study. *Thromb Res*. Mar 2018;163:41-46. doi:10.1016/j.thromres.2018.01.009
33. McBane RD, 2nd, Loprinzi CL, Zemla T, et al. Extending venous thromboembolism secondary prevention with apixaban in cancer patients. The EVE trial. *J Thromb Haemost*. Jun 2024;22(6):1704-1714. doi:10.1016/j.jtha.2024.03.011

Tables and figures

Table 1: Baseline characteristics stratified by study

Demographics	**Overall (n=2245)**	**CATCH[14] (n=455)**	**CLOT[15] (n= 338)**	**Hokusai[2] (n=1046)**	**Select-D[3] (n=406)**
Mean age, years (SD)	63.4 (11.8)	60.2 (12.9)	62.4 (11.7)	64.0 (11.3)	66.2 (10.6)
Male sex, n (%)	1102 (49.1)	189 (41.5)	159 (47.0)	540 (51.7)	214 (52.7)
Mean weight, kg (SD)	75.6 (18.0)	67.2 (17.2)	73.6 (15.5)	78.9 (18.0)	78.4 (17.4)
ECOG performance score, n (%)*					
0	591 (26.5)	88 (19.4)	80 (23.7)	303 (29.2)	120 (30.0)
1	1066 (47.8)	257 (56.6)	135 (39.9)	489 (47.1)	185 (46.2)
2	569 (25.5)	109 (24.0)	118 (34.9)	247 (23.8)	95 (23.8)
3	5 (0.2)	0 (0.0)	5 (1.5)	0 (0.0)	0 (0.0)
Li cancer type risk classification, n (%)**					
Very high-risk	298 (13.3)	60 (13.2)	18 (5.3)	143 (13.7)	77 (19.1)
High-risk	691 (30.8)	142 (31.2)	79 (23.4)	362 (34.6)	108 (26.7)
Intermediate-risk	385 (17.2)	68 (14.9)	52 (15.4)	162 (15.5)	103 (25.5)
Low-risk	867 (38.6)	185 (40.7)	187 (55.3)	379 (36.2)	116 (28.7)
Hematological cancer, n (%)	226 (10.1)	44 (9.7)	38 (11.2)	111 (10.6)	33 (8.2)

Metastatic disease, n (%)	1300 (58.8)	250 (54.9)	223 (66.0)	595 (58.2)	232 (58.6)
Use of antiplatelets, n (%)	177 (8.0)	46 (10.1)	54 (16.0)	44 (4.3)	33 (8.1)
Platelet count >350 $x10^9$/L, n (%)	371 (16.6)	102 (22.6)	73 (22.0)	126 (12.1)	70 (17.2)
Index event, n (%)					
PE ± DVT	1209 (54%)	195 (42.9%)	103 (30.5%)	657 (62.8%)	295 (72.6%)
DVT only	1036 (46%)	257 (56.0%)	235 (69.5%)	389 (37.2%)	111 (27.4%)
VTE treatment, n					
Edoxaban	522 (23.3)	0	0	522	0
Rivaroxaban	203 (9.0)	0	0	0	203
Dalteparin	1065 (47.4)	0	338	524	203
Tinzaparin	455 (20.3)	455	0	0	0
Recurrent VTE on treatment, n (%)	150 (6.7)	31 (6.8)	27 (8.0)	66 (6.3)	26 (6.4)
Recurrent VTE type, n (%)					
PE ± DVT	81 (54.0)	20 (64.5)	13 (48.1)	35 (53.0)	13 (50.0)
DVT	67 (44.7)	11 (35.5)	14 (51.9)	31 (47.0)	11 (42.3)
Other	2 (<0.1)	0 (0.0)	0 (0.0)	0 (0.0)	2 (7.7)
All-cause mortality	925 (30.4%)	150 (33.4%)	130 (38.5%)	267 (25.5%)	104 (25.6%)

* 14 patients had missing data on ECOG performance status score

** Very high-risk cancer types: pancreatic, gastroesophageal, bile duct, and gall bladder cancer; high-risk cancer types: lung, ovarian, uterine, bladder, kidney, testicular, primary brain cancer, aggressive non-Hodgkin lymphoma, multiple myeloma, and soft tissue sarcoma; intermediate-risk cancer type: colorectal cancer; low-risk cancer are all other cancer types. For 2 patients in the CLOT and 2 patients in the SELECT-D trial data on cancer type was missing.

Abbreviations: SD, standard deviation; ECOG, Eastern Cooperative Oncology Group. VTE, venous thromboembolism; DVT, deep vein thrombosis; PE, Pulmonary embolism.

Table 2. Univariable and multivariable odds ratios for prediction of on treatment recurrent VTE.

Model to predict on treatment recurrent VTE	Univariable odds ratio (95% CI)	Multivariable odds ratio (95% CI)	P-value multivariable odds ratios
Age 1 (restricted cubic spline)	0.98 (0.96-1.01)	0.99 (0.96-1.01)	0.22
Age 2 (restricted cubic spline)	0.98 (0.95-1.02)	0.98 (0.95-1.02)	0.31
Presence of metastasis	1.40 (0.85-2.30)	1.44 (1.01-2.05)	0.05
Breast cancer	0.41 (0.20-0.84)	0.42 (0.20-0.87)	0.02
Treatment with a DOAC	0.57 (0.38-0.85)	0.66 (0.44-0.98)	0.04
Index event is DVT only	1.80 (1.29-2.52)	1.72 (1.31-2.27)	<0.01

Other candidate predictors excluded during backward selection			
ECOG performance score 1 or 2	1.23 (0.83-1.83)	n.a.	n.a.
Male sex	1.13 (0.81-1.58)	n.a.	n.a.
Use of antiplatelets	0.80 (0.37-1.47)	n.a.	n.a.
Platelet count > 350 x10^9/L	0.98 (0.62-1.54)	n.a.	n.a.
Weight in kg	1.01 (0.97-1.01)	n.a.	n.a.
Lung cancer	0.99 (0.60-1.62)	n.a.	n.a.
Hepatobiliary cancer	1.53 (0.89-2.63)	n.a.	n.a.
Gynecological cancer	1.39 (0.89-2.17)	n.a.	n.a.
Urogenital cancer excluding prostate cancer	1.29 (0.68-2.45)	n.a.	n.a.
Hematological cancer	0.76 (0.41-1.40)	n.a.	n.a.
Li cancer risk classification (reference = low risk)			
Very high risk	1.47 (0.90-2.40)	n.a.	n.a.
High risk	1.12 (0.75-1.68)	n.a.	n.a.
Intermediate risk	1.02 (0.62-1.68)	n.a.	n.a.

Abbreviations: ECOG, Eastern Cooperative Oncology Group; DOAC, direct oral anticoagulant; DVT, deep vein thrombosis; CI, confidence interval; VTE, venous thromboembolism

Figure 1. C-statics and prediction interval in internal-external cross validation

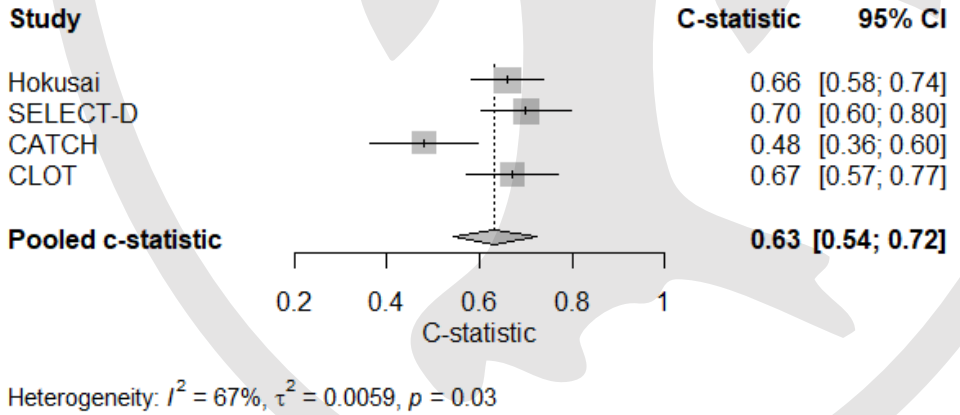
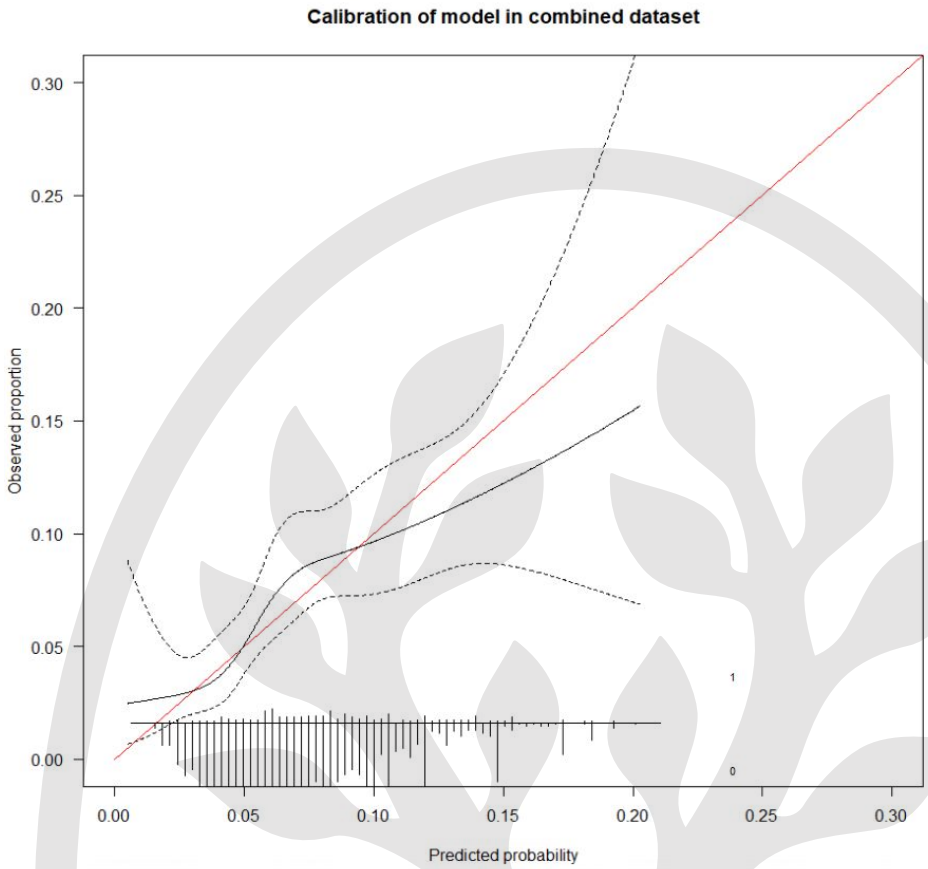


Figure 2. Calibration plot



Calibration in one imputed datasets is shown.

Supplementary Table 1. TRIPOD criteria checklist

Section/Topic	Item	Checklist Items	
Title and abstract			**Page**
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	6
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	7
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7
	5b	Describe eligibility criteria for participants.	7
	5c	Give details of treatments received, if relevant.	7
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7
	6b	Report any actions to blind assessment of the outcome to be predicted.	7
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	8
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	8
Sample size	8	Explain how the study size was arrived at.	7
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	8
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	9
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	8-9
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	9
Risk groups	11	Provide details on how risk groups were created, if done.	8-9
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	10
Model development	14a	Specify the number of participants and outcome events in each analysis.	10
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	11
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	11
	15b	Explain how to the use the prediction model.	11
Model performance	16	Report performance measures (with CIs) for the prediction model.	11
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	13-14
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	11-14
Implications	20	Discuss the potential clinical use of the model and implications for future research.	14
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	N.A
Funding	22	Give the source of funding and the role of the funders for the present study.	1

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

Supplementary Table 2. Studies eligible for inclusion

Study	Study period	Intervention	Control	Efficacy outcome	Follow-up	Patients (n)	Recurrences, n (%)	Lost to follow-up
CLOT	1999 - 2001	Dalteparin 200 IU/kg od	VKA + dakteparin for first 5-7 day	Symptomatic VTE	6 months	676	80 (11.8%)	Not reported
CATCH	2010 - 2013	Tinzaparin 175 IU/kg od	Warfarin + tinzaparin 175 IU/kg od for first 5-10 days	Symptomatic VTE	6 months	914	76 (8.4%)	14
Hokusai-VTE Cancer	2015 - 2016	Edoxaban 60 mg or 30 mg od	Dalteparin 200 IU/kg od for first 30 days followed by 150 IU/kg od	Symptomatic or incidental VTE	12 months	1,046	80 (7.6%)[†]	8
SELECT-D	2013 - 2016	Rivaroxaban 15 mg bif for first 21 days followed by 20 mg od	Dalteparin 200 IU/kg od for first 30 days followed by 150 IU/kg od	Symptomatic or incidental VTE	6 months	406	26 9 (6.4%)	1
TOTAL						**3.042**	**262 (8.6%)**	**23**

Abbreviations: VKA vitamin K antagonist; VTE, venous thromboembolism; LMWH Low-molecular-weight heparin. [†]number of events presented are during the first 6 months of the study period.

Supplementary Table 3. Cancer types in per study and in combined dataset.

	Overall	CATCH[14]	CLOT[15]	Hokusai[2]	Select-D[3]
Total number of patients	2245	455	338	1046	406
Cancer type, n (%)*					
Bladder	68 (3.0)	14 (3.1)	10 (3.0)	30 (2.9)	14 (3.5)
Brain	49 (2.2)	11 (2.4)	14 (4.2)	21 (2.0)	3 (0.7)
Breast	262 (11.7)	37 (8.1)	59 (7.6)	125 (12.0)	41 (10.1)
Cervix	74 (3.3)	46 (10.1)	14 (4.2)	14 (1.3)	0 (0.0)
Colorectal	385 (17.2)	68 (14.9)	52 (15.5)	162 (15.5)	103 (25.5)
Endometrium	55 (2.5)	18 (4.0)	0 (0.0)	37 (3.5)	0 (0.0)
Gallbladder	10 (0.4)	6 (1.3)	0 (0.0)	0 (0.0)	4 (1.0)

Gastro-esophageal	130 (5.8)	29 (6.4)	6 (1.8)	54 (5.2)	41 (10.1)
Head and neck	23 (1.0)	5 (1.1)	0 (0.0)	18 (1.7)	0 (0.0)
Hepatobiliary	37 (1.7)	9 (2.0)	0 (0.0)	26 (2.5)	2 (0.5)
Leukemia	37 (1.7)	4 (0.9)	8 (2.4)	19 (1.8)	6 (1.5)
Lung	287 (12.8)	48 (10.5)	40 (11.9)	152 (14.5)	47 (11.6)
Lymphoma	118 (5.3)	26 (5.7)	26 (7.7)	44 (4.2)	22 (5.4)
Melanoma	20 (0.9)	5 (1.1)	0 (0.0)	15 (1.4)	0 (0.0)
Multiple myeloma	63 (2.8)	14 (3.1)	4 (1.2)	40 (3.8)	5 (1.2)
Ovarian	124 (5.5)	31 (6.8)	11 (3.3)	52 (5.0)	30 (7.4)
Pancreas	121 (5.4)	16 (3.5)	12 (3.6)	63 (6.0)	30 (7.4)
Prostate	129 (5.8)	21 (4.6)	25 (7.4)	62 (5.9)	21 (5.2)
Renal	36 (1.6)	3 (0.7)	0 (0.0)	26 (2.5)	7 (1.7)
Sarcoma	36 (1.6)	8 (1.8)	0 (0.0)	26 (2.5)	2 (0.5)
Testicular	28 (1.2)	13 (2.9)	0 (0.0)	15 (1.4)	0 (0.0)
Unknown primary	23 (1.0)	3 (0.7)	5 (1.5)	9 (0.9)	6 (1.5)
Other gastrointestinal	1 (0.0)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)
Other gynecological	36 (1.6)	10 (2.2)	13 (3.9)	0 (0.0)	13 (3.2)
Other solid	81 (3.6)	9 (2.0)	37 (11.0)	28 (2.7)	7 (1.7)
Other hematological	8 (0.4)	0 (0.0)	0 (0.0)	8 (0.8)	0 (0.0)
Li cancer classification, n (%)*					
Very high risk	298 (13.3)	60 (13.2)	18 (5.3)	143 (13.7)	77 (19.1)
High risk	691 (30.8)	142 (31.2)	79 (23.4)	362 (34.6)	108 (26.7)
Intermediate risk	385 (17.2)	68 (14.9)	52 (15.4)	162 (15.5)	103 (25.5)
Low risk	867 (38.6)	185 (40.7)	187 (55.3)	379 (36.2)	116 (28.7)

* For 2 patients in the CLOT and 2 patients in the SELECT-D trial data on cancer type was missing.

Supplementary Table 4. Model for prediction of 6-month risk of on-treatment recurrent VTE

Predictors	**β**	**Standard error**	**p-value**
Intercept	-1.90	0.66	<0.01
Age 1*	-0.01	0.01	0.22
Age 2*	-0.02	0.02	0.31
Metastatic disease	0.36	0.18	0.05
Breast cancer	-0.87	0.37	0.02
Treatment with a DOAC	-0.42	0.21	0.04
DVT only as index event	0.54	0.14	<0.01

* Restricted cubic splines were used with 3 knots located at age 49, 65, and 78.

Abbreviations: DOAC, direct oral anticoagulants; DVT, deep vein thrombosis

Supplementary Table 5. Crude and adjusted hazard ratios for on-treatment recurrent VTE in original studies.

Table 5a: Crude hazard ratios for 6-month risk of on-treatment recurrent VTE including all patients (including vitamin K antagonists users).

Predictor	**Hokusai**				**Select-D**				**CATCH**		
	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI
Age	0.98	0.96	1.00	0.03	0.99	0.96	1.01	0.31	0.98	0.97	1.00
Weight	1.01	0.99	1.02	0.36	1.01	0.99	1.04	0.42	0.99	0.97	1.00
Male sex	1.05	0.65	1.70	0.84	1.45	0.66	3.20	0.35	1.14	0.70	1.85
ECOG performance score (reference is score of 0)											
- ECOG 1	1.20	0.68	2.13	0.52	2.83	0.95	8.41	0.06	1.99	0.89	4.44
- ECOG 2	1.26	0.63	2.52	0.51	2.08	0.58	7.49	0.26	3.83	1.69	8.68
Use of antiplatelets	0.37	0.05	2.68	0.32	1.01	0.24	4.36	0.98	0.25	0.06	1.02
Platelets >350	1.36	0.70	2.67	0.37	0.98	0.34	2.82	0.97	0.89	0.50	1.59
Index event DVT	1.86	1.12	3.08	0.02	2.85	1.32	6.14	0.01	1.36	0.74	2.49
Metastatic Cancer	1.88	1.11	3.20	0.02	1.87	0.81	4.34	0.14	1.66	1.00	2.76
GU or GI cancer	1.33	0.82	2.15	0.25	1.11	0.51	2.41	0.79	1.34	0.83	2.18
Hepatobiliary cancers	2.03	1.01	4.06	0.05	2.71	0.94	7.85	0.07	3.07	1.51	6.24
Genitourinary	1.27	0.69	2.33	0.44	1.17	0.45	3.01	0.75	1.94	1.19	3.15

cancers											
Breast cancer	0.53	0.21	1.33	0.18	0.33	0.04	2.45	0.28	n/e	n/e	n/e
Lung cancer	1.22	0.62	2.39	0.57	0.74	0.18	3.14	0.68	1.12	0.54	2.32
Upper GI cancers	1.61	0.39	6.61	0.51	0.77	0.18	3.27	0.73	1.64	0.65	4.11
Prostate cancer	0.23	0.03	1.66	0.14	1.36	0.32	5.78	0.67	1.89	0.76	4.74
Urological cancers	1.10	0.44	2.70	0.84	1.40	0.35	5.70	0.64	1.64	0.77	3.50
Gynecological cancers	1.36	0.65	2.85	0.42	1.02	0.31	3.30	0.98	1.77	1.06	2.94
Pancreatic cancer	1.53	0.63	3.74	0.35	3.45	1.19	9.99	0.02	2.39	0.84	6.82
Li cancer classification (reference is low risk)											
- Very high risk cancer	2.79	1.36	5.74	0.01	1.13	0.41	3.11	0.81	2.35	1.22	4.52
- High risk cancer	0.58	0.31	1.09	0.09	1.74	0.60	5.06	0.31	0.68	0.39	1.19
- Intermediate risk cancer	0.55	0.26	1.17	0.12	1.83	0.63	5.36	0.27	2.64	0.80	8.75

Abbreviations: ECOG, Eastern Cooperative Oncology Group; DOAC, direct oral anticoagulant; DVT, deep vein thrombosis; GU, genitourinary; GI, gastrointestinal; HR, hazard ratio; CI, confidence interval; VTE, venous thromboembolism

Table 5b: Age and sex adjusted hazard ratios for 6-month risk of on-treatment recurrent VTE (including vitamin K antagonists users).

Predictor	**Hokusai**				**Select-D**				**CATCH**				**CLOT**			
	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value
Weight	1.00	0.99	1.02	0.69	1.01	0.98	1.04	0.64	0.99	0.97	1.00	0.14	0.99	0.98	1.00	0.20
ECOG performance score (reference is score of 0)																
- ECOG 1	1.22	0.68	2.18	0.50	3.11	1.07	9.05	0.04	2.09	0.94	4.67	0.07	1.28	0.67	2.44	0.45
- ECOG 2	1.36	0.67	2.75	0.39	2.48	0.74	8.34	0.14	3.66	1.59	8.43	<0.0001	2.03	1.10	3.73	0.02
Use of antiplatelets	0.40	0.05	2.90	0.36	1.10	0.24	5.10	0.90	0.27	0.07	1.13	0.07	1.52	0.91	2.54	0.11

Platelets >350	1.33	0.68	2.60	0.41	1.01	0.35	2.89	0.98	0.86	0.47	1.55	0.61	0.94	0.53	1.67	0.83
Index event DVT only	2.18	1.34	3.55	0.00	3.29	1.53	7.08	0.00	0.89	0.55	1.46	0.65	1.25	0.76	2.06	0.37
Metastatic cancer	1.84	1.09	3.12	0.02	1.83	0.80	4.23	0.16	1.65	0.99	2.74	0.05	3.54	1.92	6.51	<0.001
GU or GI cancer	1.40	0.82	2.40	0.21	1.00	0.44	2.28	1.00	1.39	0.85	2.29	0.19	0.86	0.50	1.46	0.58
Hepatobiliary cancers	2.06	1.02	4.15	0.04	2.73	0.97	7.74	0.06	3.17	1.56	6.41	<0.01	2.54	1.07	6.05	0.03
Genitourinary cancers	1.21	0.67	2.21	0.53	1.27	0.47	3.40	0.64	1.99	1.16	3.42	0.01	0.87	0.42	1.79	0.70
Breast cancer	0.49	0.19	1.27	0.14	0.36	0.04	2.91	0.34	n/e	n/e	n/e	n/e	0.28	0.10	0.83	0.02
Lung cancer	1.26	0.64	2.48	0.50	0.69	0.16	3.00	0.62	1.11	0.53	2.36	0.78	2.89	1.78	4.69	<0.001
Upper GI cancers	1.69	0.42	6.87	0.46	0.77	0.18	3.22	0.72	1.63	0.64	4.14	0.30	1.30	0.35	4.78	0.69
Prostate cancer	0.24	0.03	1.75	0.16	1.29	0.29	5.66	0.74	2.36	0.87	6.43	0.09	0.59	0.21	1.64	0.31
Urological cancers	0.99	0.38	2.54	0.98	1.22	0.30	4.97	0.78	1.34	0.57	3.13	0.50	1.19	0.44	3.26	0.73
Gynecological cancers	1.42	0.64	3.16	0.39	1.29	0.34	4.88	0.70	2.33	1.19	4.56	0.01	0.69	0.26	1.81	0.45
Pancreatic cancer	1.58	0.64	3.91	0.32	3.46	1.24	9.65	0.02	2.42	0.86	6.85	0.10	2.54	1.07	6.05	0.03
Li cancer classification (reference is low risk)																
- Very high risk cancer	3.27	1.61	6.63	<0.001	0.95	0.36	2.50	0.91	2.18	1.10	4.31	0.03	2.49	1.11	5.55	0.03
- High risk cancer	0.57	0.31	1.05	0.07	1.85	0.64	5.36	0.25	0.69	0.37	1.27	0.23	0.50	0.30	0.84	0.01
- Intermediate risk cancer	0.48	0.23	0.99	0.05	1.77	0.53	5.94	0.35	2.58	0.68	9.76	0.16	1.24	0.56	2.75	0.60

Abbreviations: ECOG, Eastern Cooperative Oncology Group; DOAC, direct oral anticoagulant; DVT, deep vein thrombosis; GU, genitourinary; GI, gastrointestinal; HR, hazard ratio; CI, confidence interval; VTE, venous thromboembolism; n/e, not estimable

Table 5c: Model adjusted hazard ratios for 6-month risk of on-treatment recurrent VTE based on all included patients in original studies (including vitamin K antagonists users). This model included the following predictors after backward selection: age, sex, ECOG performance score, index event DVT, metastatic disease, Li cancer classification.

Predictor	**Hokusai**				**Select-D**				**CATCH**				**CLOT**			
	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value
Age 1	0.98	0.95	1.02	0.42	1.03	0.96	1.11	0.42	0.99	0.96	1.02	0.41	0.96	0.93	1.00	0.04
Age 2	0.98	0.94	1.03	0.44	0.91	0.82	1.02	0.10	1.00	0.95	1.06	0.94	1.01	0.96	1.06	0.81
Male sex	0.84	0.51	1.39	0.49	1.64	0.70	3.85	0.25	1.11	0.64	1.91	0.72	1.78	1.12	2.84	0.01
ECOG performance score (reference is ECOG score 0)																
- ECOG 1	1.37	0.76	2.48	0.29	2.90	1.01	8.32	0.05	1.83	0.82	4.09	0.14	0.94	0.50	1.76	0.84
- ECOG 2	1.24	0.60	2.53	0.56	2.61	0.71	9.52	0.15	3.21	1.35	7.62	0.01	1.42	0.76	2.67	0.27
Index event DVT only	1.98	1.17	3.34	0.01	3.12	1.42	6.86	<0.001	1.45	0.79	2.68	0.23	1.12	0.60	2.09	0.73
Metastatic cancer	1.67	0.99	2.83	0.06	2.40	0.97	5.93	0.06	1.43	0.82	2.49	0.21	3.43	1.80	6.54	<0.001
Li cancer classification (reference is low risk)																
- Very high risk cancer	2.61	1.26	5.39	0.01	1.14	0.42	3.07	0.80	2.17	1.04	4.50	0.04	1.99	0.88	4.52	0.10
- High risk cancer	1.67	0.89	3.12	0.11	0.56	0.19	1.60	0.28	1.34	0.72	2.49	0.35	1.95	1.15	3.31	0.01
- Intermediate risk cancer	1.75	0.82	3.73	0.15	0.43	0.12	1.58	0.20	0.38	0.10	1.38	0.14	0.72	0.32	1.59	0.41

Abbreviations: ECOG, Eastern Cooperative Oncology Group; DOAC, direct oral anticoagulant; DVT, deep vein thrombosis; GU, genitourinary; GI, gastrointestinal; HR, hazard ratio; CI, confidence interval; VTE, venous thromboembolism; n/e, not estimable

Table 5d: Crude hazard ratios for 6-month risk of on-treatment recurrent VTE excluding patients using vitamin K antagonist.

Predictor	**Hokusai**				**Select-D**				**CATCH**				**CLOT**			
	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value
Age	0,98	0,96	1,00	0,03	0,99	0,96	1,01	0,31	0,98	0,95	1,00	0,07	0,97	0,94	0,99	0,01
Weight	1,01	0,99	1,02	0,36	1,01	0,99	1,04	0,42	0,99	0,96	1,01	0,33	1,01	0,99	1,03	0,37
Male sex	1,05	0,65	1,70	0,84	1,45	0,66	3,20	0,35	0,87	0,42	1,79	0,71	1,40	0,66	2,98	0,38
ECOG performance score (reference is ECOG score 0)																
- ECOG 1	1,20	0,68	2,13	0,52	2,83	0,95	8,41	0,06	1,09	0,40	2,99	0,87	0,73	0,26	1,99	0,53
- ECOG 2	1,26	0,63	2,52	0,51	2,08	0,58	7,49	0,26	2,79	1,03	7,54	0,04	1,43	0,57	3,59	0,44
Use of antiplatelets	0,37	0,05	2,68	0,32	1,01	0,24	4,36	0,98	0,28	0,04	2,12	0,22	1,22	0,47	3,18	0,68
Platelets >350	1,36	0,70	2,67	0,37	0,98	0,34	2,82	0,97	0,53	0,19	1,50	0,23	1,33	0,56	3,17	0,52
Metastatic cancer	1,88	1,11	3,20	0,02	1,87	0,81	4,34	0,14	0,95	0,47	1,93	0,89	1,73	0,73	4,06	0,21
Index event DVT only	1,86	1,12	3,08	0,02	2,85	1,32	6,14	0,01	0,98	0,44	2,19	0,96	0,96	0,36	2,51	0,93
GU or GI cancer	1,33	0,82	2,15	0,25	1,11	0,51	2,41	0,79	1,26	0,62	2,56	0,52	0,90	0,40	2,05	0,80
Hepatobiliary cancers	2.03	1.01	4.06	0.05	2.71	0.94	7.85	0.07	1.79	0.53	6.01	0.35	n/e	n/e	n/e	n/e
Genitourinary cancers	1.27	0.69	2.33	0.44	1.17	0.45	3.01	0.75	2.26	1.12	4.57	0.02	0.78	0.23	2.58	0.68
Breast cancer	0.53	0.21	1.33	0.18	0.33	0.04	2.45	0.28	n/e	n/e	n/e	n/e	0.36	0.09	1.50	0.16
Lung cancer	1.22	0.62	2.39	0.57	0.74	0.18	3.14	0.68	0.63	0.15	2.59	0.52	1.92	0.73	5.06	0.19
Upper GI cancers	1.61	0.39	6.61	0.51	0.77	0.18	3.27	0.73	2.26	0.69	7.41	0.18	n/e	n/e	n/e	n/e
Urological	1.	0.4	2.7	0.8	1.	0.3	5.7	0.6	1.	0.4	4.4	0.5	1.	0.1	8.3	0.8

cancers	10	4	0	4	40	5	0	4	40	4	1	7	17	7	1	7
Gynecological cancers	1.36	0.65	2.85	0.42	1.02	0.31	3.30	0.98	2.21	1.08	4.54	0.03	0.67	0.16	2.85	0.59
Pancreatic cancer	1.53	0.63	3.74	0.35	3.45	1.19	9.99	0.02	2.58	0.58	11.60	0.22	n/e	n/e	n/e	n/e
Li cancer classification (reference is low risk)																
- Very high risk	2,79	1,36	5,74	0,01	1,13	0,41	3,11	0,81	1,65	0,64	4,24	0,30	0,00	0,00	0,00	0,00
- High risk	0,58	0,31	1,09	0,09	1,74	0,60	5,06	0,31	1,05	0,47	2,37	0,90	0,63	0,26	1,52	0,31
- Intermediate risk	0,55	0,26	1,17	0,12	1,83	0,63	5,36	0,27	5,52	0,73	42,00	0,10	0,60	0,23	1,58	0,30

Abbreviations: ECOG, Eastern Cooperative Oncology Group; DVT, deep vein thrombosis; GU, genitourinary; GI, gastrointestinal; HR, hazard ratio; CI, confidence interval; VTE, venous thromboembolism; n/e, not estimable.

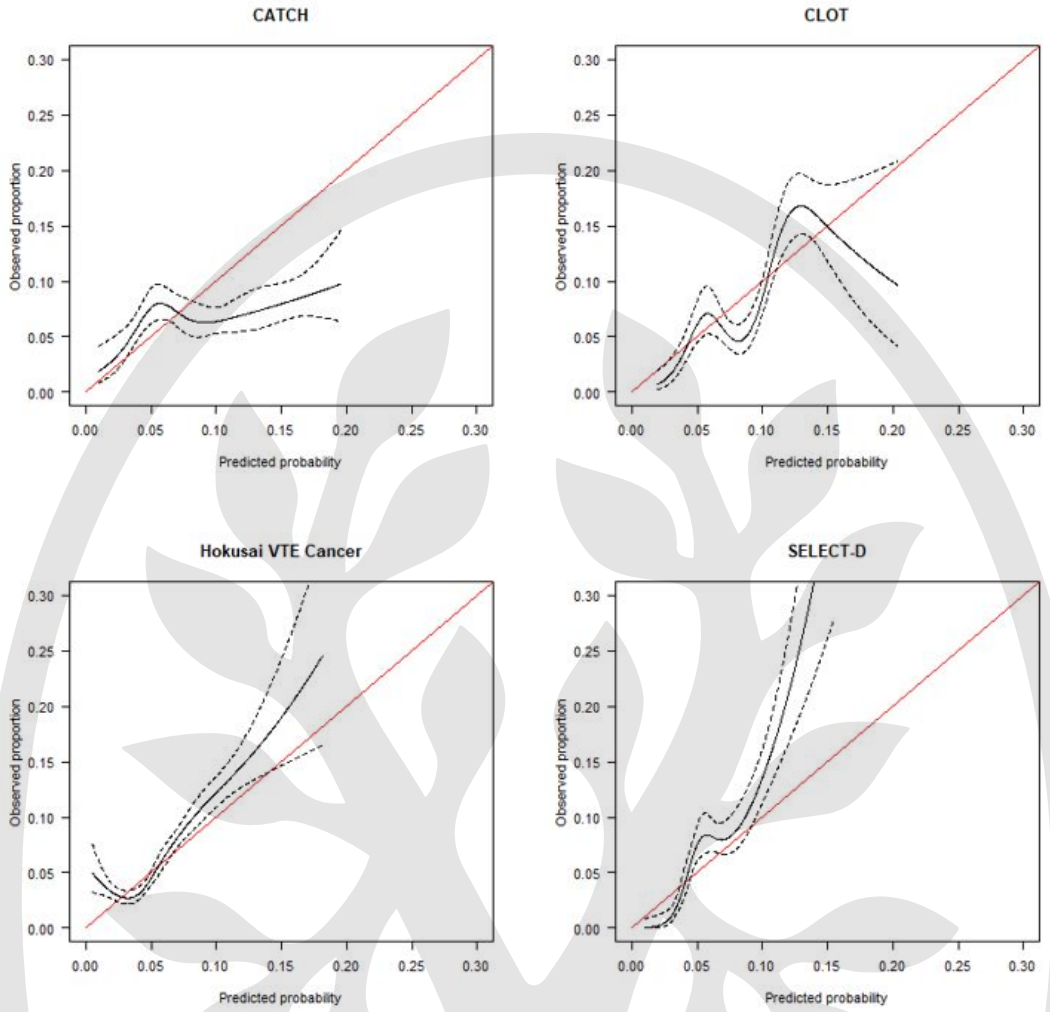
Table 5e: Age and sex adjusted hazard ratios for 6-month risk of on-treatment recurrent VTE excluding patients using vitamin K antagonist.

Predictor	**Hokusai**				**Select-D**				**CATCH**				**CLOT**			
	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value	HR	lower CI	upper CI	p-value
Weight	1,000	0,99	1,02	0,69	1,001	0,98	1,04	0,64	0,99	0,97	1,02	0,47	1,000	0,99	1,02	0,58
ECOG performance score (reference is ECOG score 0)																
- ECOG 1	1,22	0,68	2,18	0,50	3,111	1,07	9,05	0,04	1,14	0,42	3,08	0,80	0,883	0,29	2,42	0,74
- ECOG 2	1,36	0,67	2,75	0,39	2,48	0,74	8,34	0,14	2,63	0,93	7,42	0,07	1,666	0,63	4,41	0,31
Use of antiplatelets	0,40	0,05	2,90	0,36	1,10	0,24	5,10	0,90	0,33	0,05	2,41	0,28	1,32	0,51	3,44	0,57
Platelets >350	1,33	0,68	2,60	0,41	1,001	0,35	2,89	0,98	0,47	0,16	1,36	0,16	1,22	0,51	2,97	0,65
Index event DVT only	2,18	1,34	3,55	0,00	3,29	1,53	7,08	0,00	0,69	0,34	1,41	0,31	1,45	0,60	3,53	0,41
Metastatic cancer	1,84	1,09	3,12	0,02	1,830	0,80	4,23	0,16	0,92	0,45	1,87	0,81	2,05	0,90	4,67	0,09
GU or GI	1,	0,8	2,4	0,2	1,	0,4	2,2	1,0	1,	0,7	2,8	0,2	0,	0,3	2,3	0,8

cancer	40	2	0	1	00	4	8	0	44	4	3	8	92	6	5	7
Hepatobiliary cancers	2.06	1.02	4.15	0.04	2.73	0.97	7.74	0.06	1.74	0.51	5.91	0.37	n/e	n/e	n/e	n/e
Genitourinary cancers	1.21	0.67	2.21	0.53	1.27	0.47	3.40	0.64	2.04	0.96	4.34	0.06	0.85	0.24	3.04	0.81
Breast cancer	0.49	0.19	1.27	0.14	0.36	0.04	2.91	0.34	n/e	n/e	n/e	n/e	0.38	0.07	1.93	0.24
Lung cancer	1.26	0.64	2.48	0.50	0.69	0.16	3.00	0.62	0.62	0.14	2.70	0.52	1.88	0.71	4.98	0.20
Upper GI cancers	1.69	0.42	6.87	0.46	0.77	0.18	3.22	0.72	2.28	0.70	7.41	0.17	n/e	n/e	n/e	n/e
Prostate cancer	0.24	0.03	1.75	0.16	1.29	0.29	5.66	0.74	6.08	1.79	20.72	0.00	0.42	0.06	3.14	0.40
Urological cancers	0.99	0.38	2.54	0.98	1.22	0.30	4.97	0.78	1.17	0.36	3.80	0.79	1.16	0.17	7.72	0.88
Gynecological cancers	1.42	0.64	3.16	0.39	1.29	0.34	4.88	0.70	2.55	0.99	6.55	0.05	0.74	0.16	3.51	0.70
Pancreatic cancer	1.58	0.64	3.91	0.32	3.46	1.24	9.65	0.02	2.50	0.56	11.18	0.23	n/e	n/e	n/e	n/e
Li cancer classification (reference is low risk)																
- Very high risk	3.27	1.61	6.63	0.00	0.95	0.36	2.50	0.91	1.61	0.60	4.28	0.34	n/e	n/e	n/e	n/e
- High risk	0.57	0.31	1.05	0.07	1.85	0.64	5.36	0.25	1.08	0.44	2.65	0.87	0.64	0.26	1.59	0.33
- Intermediate risk	0.48	0.23	0.99	0.05	1.77	0.53	5.94	0.35	5.52	0.70	43.23	0.10	0.56	0.19	1.63	0.29

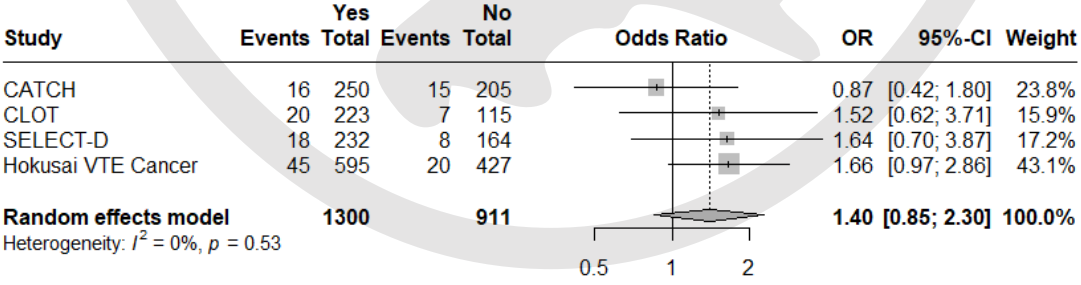
Abbreviations: ECOG, Eastern Cooperative Oncology Group; DVT, deep vein thrombosis; GU, genitourinary; GI, gastrointestinal; HR, hazard ratio; CI, confidence interval; VTE, venous thromboembolism; n/e, not estimable

Supplementary Figure 1. Calibration of the model in each study

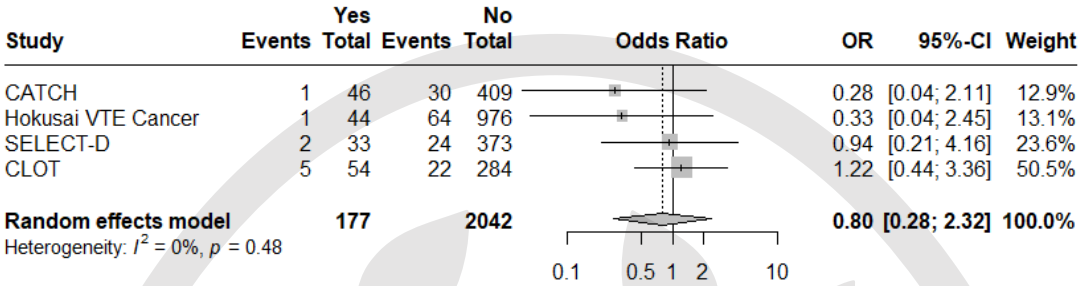


Supplementary Figure 2. Forest plots of relevant binary candidate predictors.

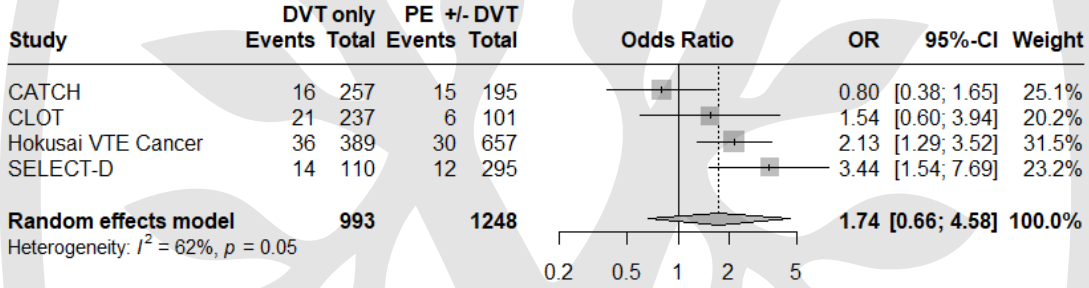
Metastatic disease yes or no



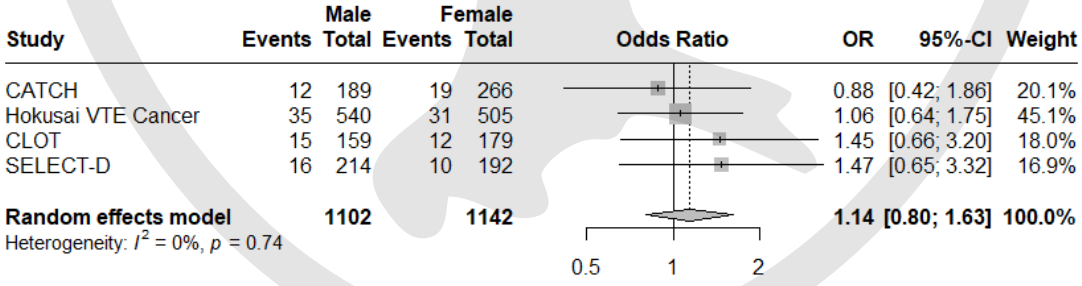
Use of antiplatelets yes or no



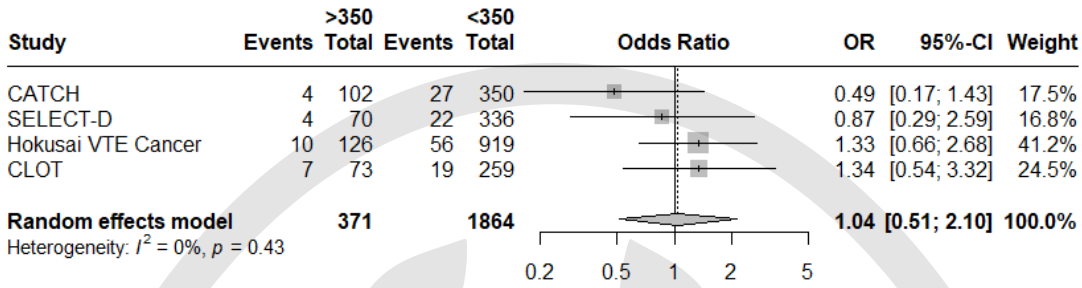
Index event deep vein thrombosis only



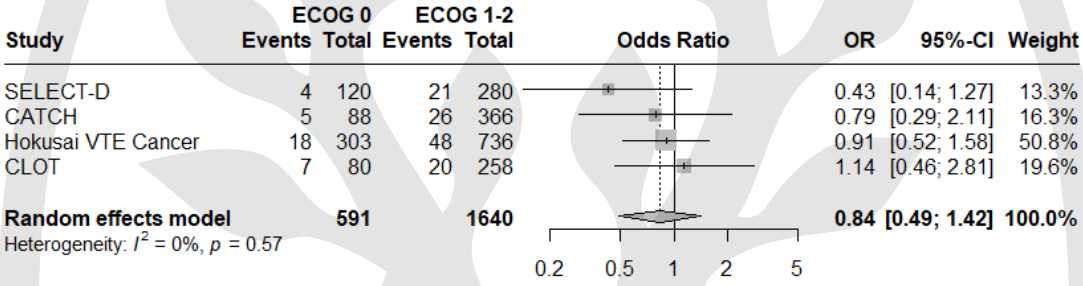
Sex



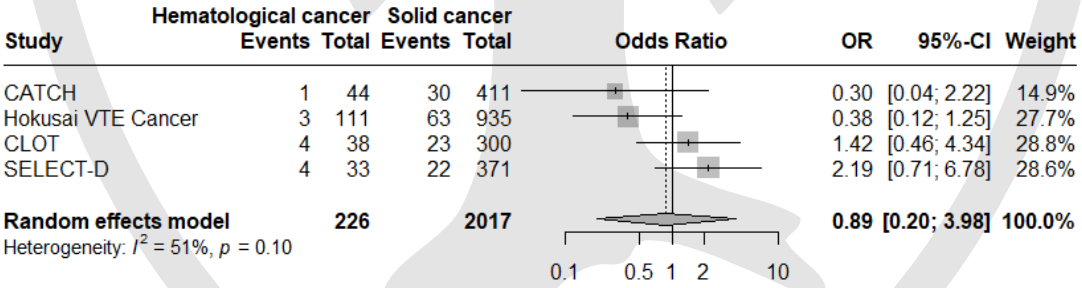
Platelet count >350 x10^9/L



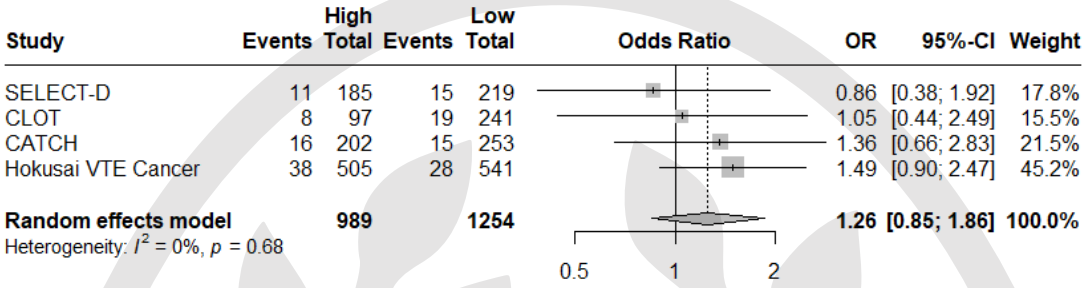
Eastern Cooperative Oncology Group score 0 vs 1-2



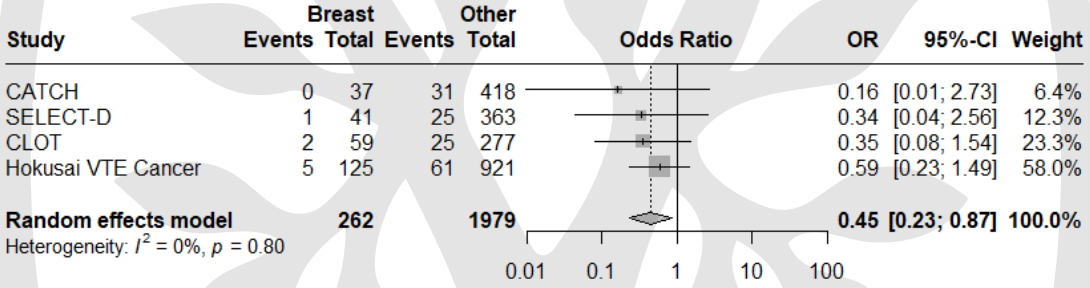
Hematological cancer vs solid cancer



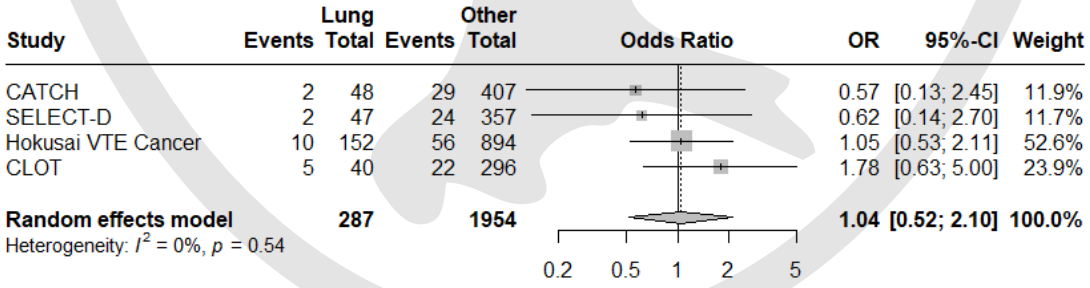
Li cancer classification high and very high risk cancer types vs Intermediate and low risk cancer types



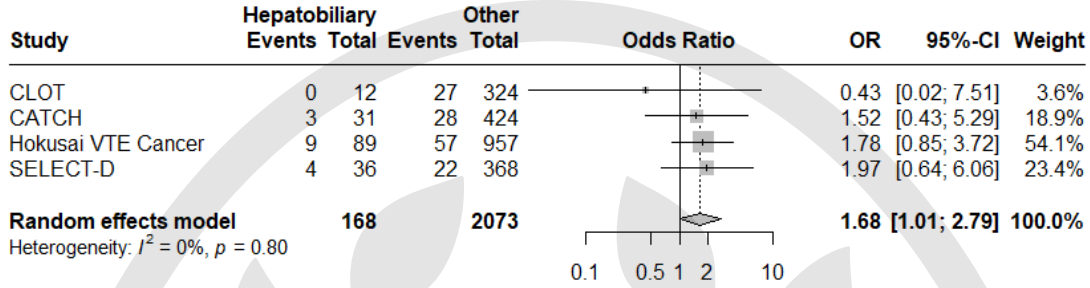
Breast cancer



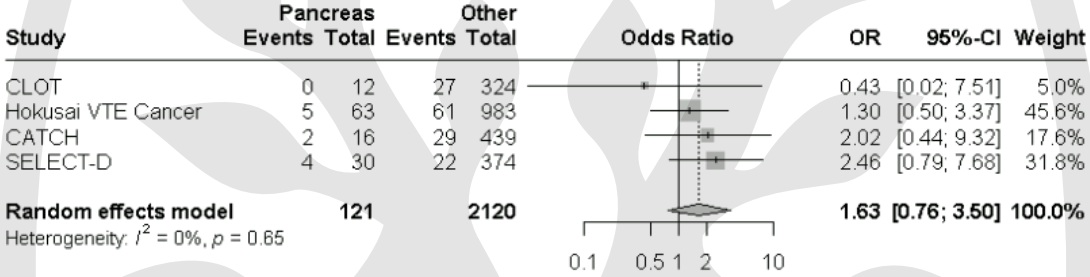
Lung cancer



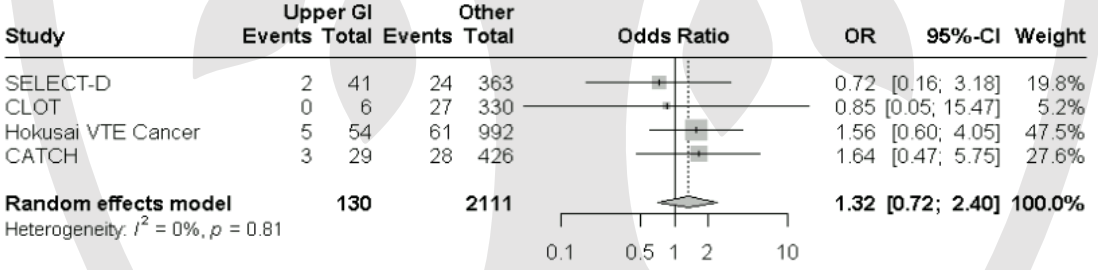
Hepatobiliary cancer



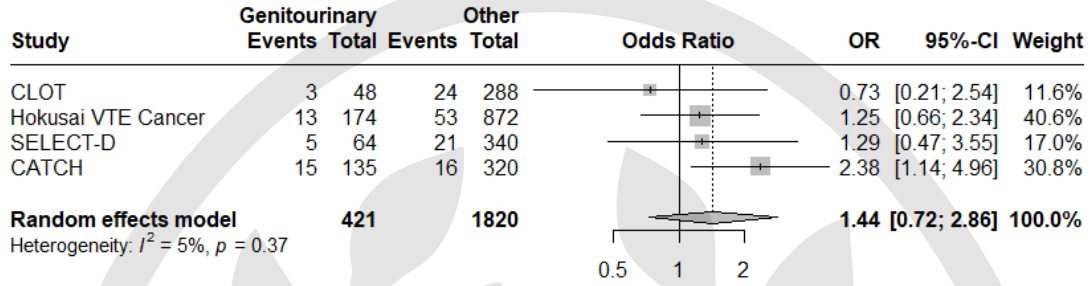
Pancreatic cancer



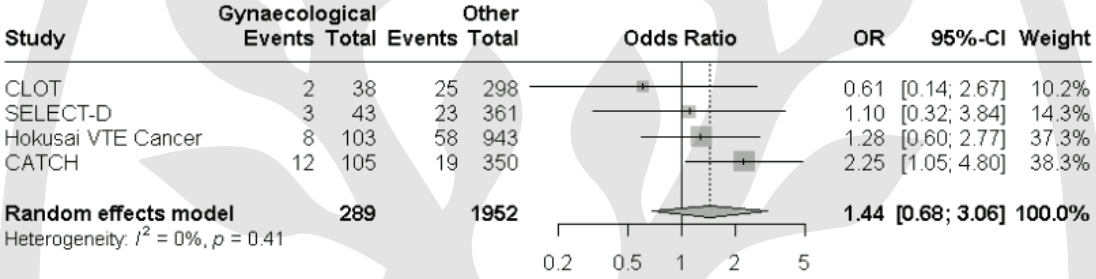
Upper gastrointestinal cancer



Genitourinary cancer excluding prostate cancer



Gynaecological cancer



Use of LMWH (CLOT and CATCH vs VKA; Hokusai and Select-D vs DOAC)