# A clinical prediction model for cancer-associated venous thromboembolism: a development and validation study in two independent prospective cohorts

*Ingrid Pabinger, Nick van Es, Georg Heinze, Florian Posch, Julia Riedl, Eva-Maria Reitter, Marcello Di Nisio, Gabriela Cesarman-Maus, Noémie Kraaijpoel, Christoph Carl Zielinski, Harry Roger Büller, Cihan Ay*

## Summary

**Background** Venous thromboembolism is a common complication of cancer, but the risk of developing venous thromboembolism varies greatly among individuals and depends on numerous factors, including type of cancer. We aimed to develop and externally validate a clinical prediction model for cancer-associated venous thromboembolism.

**Methods** We used data from the prospective Vienna Cancer and Thrombosis Study (CATS) cohort (n=1423) to select prognostic variables for inclusion in the model. We then validated the model in the prospective Multinational Cohort Study to Identify Cancer Patients at High Risk of Venous Thromboembolism (MICA) cohort (n=832). We calculated c-indices to show how the predicted incidence of objectively confirmed venous thromboembolism at 6 months compared with the cumulative 6-month incidences observed in both cohorts.

**Findings** Two variables were selected for inclusion in the final clinical prediction model: tumour-site risk category (low or intermediate *vs* high *vs* very high) and continuous D-dimer concentrations. The multivariable subdistribution hazard ratios were 1·96 (95% CI 1·41–2·72; p=0·0001) for high or very high versus low or intermediate and 1·32 (95% CI 1·12–1·56; p=0·001) per doubling of D-dimer concentration. The cross-validated c-indices of the final model were 0·66 (95% CI 0·63–0·67) in CATS and 0·68 (0·62–0·74) in MICA. The clinical prediction model was adequately calibrated in both cohorts.

**Interpretation** An externally validated clinical prediction model incorporating only one clinical factor (tumour-site category) and one biomarker (D-dimer) predicted the risk of venous thromboembolism in ambulatory patients with solid cancers. This simple model is a considerable improvement on previous models for predicting cancer-associated venous thromboembolism, and could aid physicians in selection of patients who will likely benefit from thromboprophylaxis.

**Funding** Austrian Science Fund, Austrian National Bank Memorial Fund, and participating hospitals.



## Introduction

Venous thromboembolism, including deep vein thrombosis and pulmonary embolism, is a common complication of cancer, with an incidence of 1–20%.[1] Randomised trials have shown that prophylactic anticoagulation with low-molecular-weight heparin approximately halves the relative risk of venous thromboembolism in patients with cancer.[2] However, the absolute risk reduction of this intervention appears to be modest for most ambulatory patients with cancer, in whom the risk of venous thromboembolism is about 3–5% in the first months of chemotherapy.[2] Moreover, bleeding complications associated with anticoagulant therapy are common in patients with cancer,[3] and in patients at very low risk of venous thromboembolism the potential harms of thrombo-prophylaxis might exceed the benefits because of an increased risk of bleeding.[2] Therefore, the decision to provide anticoagulation for prevention of cancer-associated venous thromboembolism should ideally be informed by a valid risk-stratification strategy.[4] With this personalised approach, thromboprophylaxis could be provided to those patients at greatest risk of developing venous thromboembolism and avoided in low-risk patients.

The most widely used clinical prediction model for this purpose is the Khorana score, which aims to identify ambulatory patients with cancer at increased risk of venous thromboembolism during chemotherapy by use of two clinical variables (tumour site and body-mass index) and three laboratory measurements (platelets, haemoglobin, and leucocytes).[5,6] Other scores use different variables, such as the Vienna modification of the Khorana score (addition of biomarkers D-dimer and soluble P-selectin),[7] the PROTECHT score (addition of gemcitabine and platinum-based chemotherapy),[8] and the CONKO score (addition of WHO performance status).[9] However, in a prospective validation study[10]—of 616 patients on chemotherapy and 260 patients who had not yet received chemotherapy—only two scoring approaches, the Vienna modification and the PROTECHT

*Lancet Haematol* 2018; 5: e289–98

Published Online June 6, 2018 http://dx.doi.org/10.1016/S2352-3026(18)30063-2

**This online publication has been corrected. The corrected version first appeared at thelancet.com/haematology on June 15, 2018**

See Comment page e273

**Clinical Division of Haematology and Haemostaseology** (Prof I Pabinger MD, F Posch MD, J Riedl MD, E-M Reitter MD, C Ay MD) **and Clinical Division of Oncology** (Prof C C Zielinski MD), **Department of Medicine I, and Section for Clinical Biometrics, Centre for Medical Statistics, Informatics and Intelligent Systems** (G Heinze PhD), **Medical University of Vienna, Vienna, Austria; Department of Vascular Medicine, Academic Medical Centre Amsterdam, University of Amsterdam, Amsterdam, Netherlands** (N van Es MD, N Kraaijpoel MD, Prof H R Büller MD)**; Division of Oncology, Department of Internal Medicine, Medical University of Graz, Graz, Austria** (F Posch)**; Department of Medicine and Ageing Sciences, G D'Annunzio University, Chieti, Italy** (M Di Nisio MD)**; and Department of Hematology, National Cancer Institute Mexico, Mexico City, Mexico** (G Cesarman-Maus MD)

Correspondence to: Prof Ingrid Pabinger, Clinical Division of Haematology and Haemostaseology, Department of Medicine I, Medical University of Vienna, Vienna 1090, Austria **ingrid.pabinger@meduniwien.ac.at**

## Research in context

### Evidence before this study

Venous thromboembolism is a common complication in patients with cancer. Although pharmacological thromboprophylaxis significantly reduces the relative risk of cancer-associated venous thromboembolism, this intervention has not been routinely adopted in clinical practice because the absolute risk reduction is low for most patients. A personalised approach to risk assessment for venous thromboembolism in the oncological setting using clinical prediction models might help clinicians to identify patients at high risk of this complication, thus justifying pharmacological thromboprophylaxis. Clinical parameters and biomarkers of haemostatic activation, such as D-dimer, thrombin, and soluble P-selectin, have previously been associated with venous thromboembolism risk in patients with cancer. Several risk scores for cancer-associated venous thromboembolism have been developed, including the Khorana score, the PROTECHT score, the Vienna update of the Khorana score, and the CONKO score. However, the performance of these scores is limited by inadequate predictive power and poor usability.

### Added value of this study

We developed and externally validated a clinical prediction model for cancer-associated venous thromboembolism in ambulatory patients with solid tumors. We considered numerous clinical variables and biomarkers during model development, but only included two in the final model: tumour-site category (the most important component of the Khorana score) and D-dimer concentrations. We showed that our clinical prediction model could outperform previous clinical prediction scores in predicting those patients at high risk of developing venous thromboembolism. The model is available for clinical use as a printed nomogram and as an online prediction tool.

### Implications of all the available evidence

Our simple clinical prediction model considerably improved prediction of cancer-associated venous thromboembolism, and could aid physicians in selection of those ambulatory patients with solid tumours who will most benefit from pharmacological thromboprophylaxis.

score, were able to accurately predict the development of venous thromboembolism. A more recently described score (COMPASS-CAT), which includes cancer-related and treatment-related factors, had good discriminatory capacity, but it was only tested in breast, colon, lung, and ovarian cancers and has not yet been externally validated.[11]

The risk of venous thromboembolism varies according to tumour site.[1] Biomarkers that reflect activation of the haemostatic system—such as D-dimer, thrombin generation, and soluble P-selectin—are independent prognostic factors for venous thromboembolism in patients with cancer.[7,12] These markers can facilitate the clinical prediction of cancer-associated venous thromboembolism.[7] However, tests for thrombin generation and soluble P-selectin are rarely available in routine clinical practice, and so they have not been included in clinical prediction models.[12]

In this study, we aimed to address some of these issues through the development and external validation of a clinical prediction model for venous thromboembolism in ambulatory patients with active solid cancers. We aimed to design a simple model to predict risk of venous thromboembolism over 6 months that could easily be used in routine clinical practice, to allow targeted thromboprophylaxis in patients at high risk of venous thromboembolism.

## Methods

### Study design and participants

See Online for appendix

We used data from two independent prospective cohorts to develop and externally validate a clinical model to predict venous thromboembolism. Both cohorts were started to identify risk factors for venous thromboembolism in people with cancer.

For model development, we used data from 1737 patients in the Vienna Cancer and Thrombosis Study (CATS)[13] who had a solid cancer (excluding primary brain tumours) or lymphoma, if classifiable by Ann Arbor staging. CATS is an ongoing, prospective, single-centre, observational cohort study with a baseline biobank. Patients with a newly diagnosed active cancer, or patients who had disease progression after complete or partial remission, were enrolled at a single tertiary academic centre in Vienna, Austria, between Oct 14, 2003, and March 26, 2014. Detailed inclusion and exclusion criteria have been described previously.[7,14] Briefly, eligible patients were older than 18 years and had a histologically confirmed cancer diagnosis. Exclusion criteria were overt bacterial or viral infection within the past 2 weeks, venous or arterial thromboembolism within the past 3 months, and ongoing treatment with continuous or direct anticoagulants (vitamin K antagonists or low-molecular-weight heparin). Patients were allowed to take aspirin, ticlopidine, or clopidogrel, and immobilised patients were treated with low-molecular-weight heparin to prevent thrombosis during hospital stay. Other exclusion criteria were surgery or radiotherapy within the past 2 weeks and chemotherapy within the past 3 months, to exclude a transient effect of these interventions on the haemostatic system. Patients were followed up until venous thromboembolism, death, or censoring at 24 months. D-dimer concentrations in blood samples collected at baseline were measured with the STA-Liatest assay (Diagnostica-Stago, Asnières, France). Other laboratory assays done in CATS are reported in appendix pp 2, 3).

The primary outcome of CATS was symptomatic, objectively confirmed, and independently assessed venous thromboembolism, defined as a composite of distal or proximal deep vein thrombosis of the leg, upper

limb deep vein thrombosis, symptomatic splanchnic deep vein thrombosis, or pulmonary embolism, occurring during a 2-year observation period. Pulmonary embolisms found incidentally were counted as events because the adjudication committee deemed them to be clinically relevant, with a requirement for anticoagulation. Upper limb deep vein thromboses related to indwelling venous catheters and incidental splanchnic vein thromboses were not considered as events.

For model validation, we used demographic, laboratory, and outcome data from the Multinational Cohort Study to Identify Cancer Patients at High Risk of Venous Thromboembolism (MICA).[10] MICA is a completed prospective, multinational, observational cohort study in 1027 patients with advanced solid cancer who were enrolled from seven centres in the Netherlands, France, Italy, and Mexico between July 27, 2008, and Feb 25, 2016. Ambulatory patients with lung, oesophageal, colorectal, pancreatic, breast, prostate, gastric, ovarian, or bladder cancer were eligible if they were scheduled for chemotherapy within 7 days of study entry or had started chemotherapy in the previous 3 months. Exclusion criteria were prophylactic or therapeutic anticoagulation or adjuvant therapy. Patients were followed up for a maximum of 6 months, until the occurrence of venous thromboembolism, death, censoring because of curative surgery (only for patients receiving neoadjuvant therapy), initiation of anticoagulation for other reasons, or loss to follow-up. D-dimer concentrations in blood samples collected at baseline were measured with the INNOVANCE assay (Siemens Healthcare, Marburg, Germany).

The primary outcome in MICA was a composite of objectively confirmed symptomatic or incidental pulmonary embolism, distal or proximal deep vein thrombosis, non-catheter-related upper limb deep vein thrombosis, or symptomatic catheter-related upper limb deep vein thrombosis, occurring during 6 months of follow-up. All diagnoses were centrally verified on the basis of imaging results. Asymptomatic upper limb deep vein thrombosis related to indwelling venous catheters and splanchnic vein thrombosis were not considered as events. Routine screening for venous thromboembolism was not done in CATS or MICA. External validation was done in a semi-masked manner without data pooling—ie, the MICA investigators had no access to CATS data and the CATS investigators had no access to MICA data.

Both studies were approved by the ethics committees of each of the participating hospitals. All patients provided written informed consent. This report adheres to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines.[15]

### Clinical prediction model development

To develop the clinical prediction model for venous thromboembolism in patients with cancer, we prespecified five clinical principles. First, we chose ambulatory patients with cancer as the target population because about 75% of all cases of cancer-associated venous thromboembolism occur within this population.[4] Second, we excluded patients with high-grade gliomas or multiple myelomas because specific clinical models to predict venous thromboembolism already exist for these tumour types.[16,17] Third, we considered the cumulative 6-month risk of venous thromboembolism as the primary endpoint. On the one hand, an interval shorter than 6 months would only have provided guidance during the first chemotherapy cycles, and the risk of developing venous thromboembolism remains high throughout the first 6 months of treatment.[7] On the other hand, an interval longer than 6 months would have been clinically irrelevant given that few physicians prescribe primary thromboprophylaxis for more than 6 months because the risk of venous thromboembolism is highest during the first 6 months after diagnosis of cancer.[2] Fourth, tumour sites were categorised by risk of venous thromboembolism as low or intermediate risk, high risk, or very high risk, according to the modified Khorana score criteria.[5,7] Because the risk of venous thromboembolism in patients with colorectal cancer was substantial in CATS (8%),[7] this type of cancer was assigned to the high-risk group. All risk group assignments were made before the development of our model. Fifth, prognostic variables were selected from a large pool of clinical and laboratory candidate variables in CATS.

### Statistical analysis

All statistical analyses were done with R version 3.3.3. We assessed distributional differences in baseline variables between CATS and MICA using measures of standardised mean differences (SMDs; values >0·2 were considered to indicate a potentially relevant difference between the two cohorts).[18] The cumulative incidence of venous thromboembolism in the two cohorts was estimated with cause-specific cumulative incidence estimators, treating death not related to venous thromboembolism as a competing event.[19] Likewise, cumulative incidence of all-cause mortality was estimated treating venous thromboembolism as a competing event.

We used a penalised regression approach (least absolute shrinkage and selection operator) to model the cause-specific risk of venous thromboembolism, with inclusion of the prognostic variables selected from CATS (appendix pp 2, 3).[20,21] Continuous variables were $\log_2$-transformed before variable selection to avoid a disproportional effect of high values. Continuous variables with standardised hazard ratios (HRs) between 0·80 and 1·25 were omitted to prevent the inclusion of variables with a small magnitude of association.

We further reduced the model by fitting a Fine and Gray competing risk regression using a backward selection algorithm with a p value greater than 0·05 indicating exclusion. Missing data in selected variables were multiply imputed with the predictive

mean-matching method in a chained equations algorithm, generating five imputed datasets. These datasets were analysed separately, and the results were pooled with Rubin's rules. The resulting model was simplified into a nomogram. External validation was done in MICA with complete-case analysis.

To assess the performance of our model, we measured discrimination (the model's ability to distinguish between patients who did and did not develop venous thromboembolism, as indicated by modification of Harrell's c-index to accommodate censoring and competing risks) using R and calibration (agreement between observed and predicted proportions of patients with venous thromboembolism) using calibration plots.[15,22] c-indices were cross-validated with 1000 bootstrap samples to account for potential over-optimism.

The final clinical prediction model was compared with the Khorana score through estimation of a population-weighted net reclassification improvement (NRI) statistic.[23] We also did a decision-curve analysis to assess the clinical usefulness of the model in indicating thromboprophylaxis compared with the approaches of universal thromboprophylaxis (a treat-all strategy) or no thromboprophylaxis (a treat-none strategy).

On the basis of our definition of a positive test (one in which the predicted 6 months' risk of venous thromboembolism exceeded a predefined cutoff), we calculated sensitivity, specificity, positive predictive value, and negative predictive value using standard formula.

CATS is registered with the Medical University Vienna (EK 126/2003) and MICA is registered with ClinicalTrials.gov (NCT02095925).

### Role of the funding source

The funders had no role in study design, data analysis, data interpretation, or writing of the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Results

Of the 1737 patients in the CATS cohort, we excluded 305 from model development because they had a primary brain tumour (n=240), unavailable D-dimer values (n=16), or lymphoma that could not be staged by Ann Arbor (n=5), because they were lost to follow-up (n=39), or because they had a secondary primary cancer identified on data review (which made assignment of the patient to a single tumour-site category impossible; n=5). Of the initial 1027 patients in MICA who were eligible for inclusion, 205 were excluded because they were receiving prophylactic or therapeutic anticoagulation (n=12) or adjuvant chemotherapy (n=56), because they had a haematological cancer (other than lymphoma; n=14), or because D-dimer testing was not done (n=123). Therefore, the development cohort (CATS) included 1423 patients and the validation cohort (MICA) included 832 patients.

As indicated by SMDs, relevant baseline variables such as age, sex, and D-dimer concentrations were similarly distributed in CATS and MICA (table 1). However, the proportion of patients with oesophageal cancer was higher in MICA than in CATS (21% *vs* 1%; SMD 0·69), and the proportion of patients with lymphoma was higher in CATS than in MICA (17% *vs* 0%; 0·65). Other differences were in the proportions of patients with newly diagnosed cancer (71% in CATS *vs* 78% in MICA; 0·17) and those receiving chemotherapy (0% in CATS *vs* 73% in MICA; 2·30). The mean Khorana score was slightly higher in MICA (1·4 points [SD 1·0]) than in CATS (1·1 points [1·0]; SMD 0·28).

During a median follow-up of 180 days (IQR 180–180 for CATS and 109–180 for MICA), 80 (6%) of 1423 patients in CATS and 48 (6%) of 832 patients in MICA developed venous thromboembolism. In a competing-risk analysis, the cumulative 6-month risk of venous thromboembolism was 5·7% (95% CI 4·5–6·9) in CATS and 6·3% (4·7–8·2) in MICA (figure 1A). The most frequent types of venous thromboembolism in CATS and MICA were lower limb deep vein thrombosis and pulmonary embolism (appendix p 6). 177 (12%) patients in CATS and 132 (16%) of patients in MICA died during follow-up. The estimated 6-month risk of mortality was 12·5% (95% CI 10·8–14·2) in CATS and 17·9% (16·2–19·7) in MICA (figure 1B).

Univariable modelling of cause-specific venous thromboembolism hazards identified 11 clinical prognostic factors and biomarkers (appendix pp 4, 5). Of these risk factors, the prespecified variable selection process selected two variables for inclusion in the clinical prediction model: tumour-site risk category (very high *vs* high and high *vs* low or intermediate) and continuous D-dimer concentrations. In this model, the multivariable sub-distribution HRs were 1·96 (95% CI 1·41–2·72; p=0·0001) for very high versus high and high versus low or intermediate risk categories and 1·32 (1·12–1·56; p=0·001) per doubling of D-dimer.

Using the model, we predicted that the mean 6-month risk of venous thromboembolism in CATS was 5·7% (range 2·2–36·0). The cross-validated c-index of this model in CATS was 0·66 (95% CI 0·63–0·67). The model was adequately calibrated (figure 2A), with no indication of systematic under-estimation or overestimation of venous thrombo-embolism in CATS. We simplified the model into a nomogram (figure 3). No significant interaction between tumour-site risk category and D-dimer concentration was observed (p=0·18), suggesting that D-dimer might be useful for further risk stratification within each tumour-site category. With the cutoff for predicted cumulative 6-month risk of venous thromboembolism in CATS set at 10%, the sensitivity of the model was 33% (95% CI 23–47), the specificity was 84% (83–87), the positive predictive value was 12% (8–16), and the negative predictive value was 95% (94–96). At a cutoff of 15%, the sensitivity of the model was 15% (8–24),

the specificity was 96% (95–97), the positive predictive value was 18% (9–29), and the negative predictive value was 95% (94–96).

Using the nomogram, we predicted that the mean 6-month risk of venous thromboembolism in the MICA cohort was 6·4% (range 2·3–23·0). The cross-validated

| | CATS (development cohort; n=1423) | | MICA (validation cohort; n=832) | | SMD* |
|---|---|---|---|---|---|
| | n† | Summary measure | n† | Summary measure | |
| Age at entry (years) | 1423 | 62·9 (54·0–68·9) | 832 | 63·7 (55·9–70·3) | 0·16 |
| Body-mass index (kg/m²) | 1418 | 25·0 (22·1–28·3) | 825 | 24·7 (22·5–27·4) | 0·09 |
| Sex | | | | | |
| Male | 1423 | 772 (54%) | 832 | 478 (57%) | 0·06 |
| Female | 1423 | 651 (46%) | 832 | 354 (43%) | 0·06 |
| Use of erythropoiesis-stimulating drugs | 1423 | 50 (4%) | 823 | 20 (2%) | 0·06 |
| Receiving chemotherapy | 1423 | 0 | 832 | 604 (73%) | 2·30 |
| Tumour-site risk category | 1423 | .. | 832 | .. | .. |
| Low or intermediate | .. | 379 (27%) | .. | 144 (17%) | 0·23 |
| Breast | .. | 226 (16%) | .. | 89 (11%) | 0·15 |
| Prostate | .. | 153 (11%) | .. | 39 (5%) | 0·23 |
| Other | .. | 0 | .. | 16 (2%) | 0·20 |
| High | .. | 863 (61%) | .. | 535 (64%) | 0·08 |
| Lung | .. | 292 (21%) | .. | 183 (22%) | 0·04 |
| Colorectal | .. | 173 (12%) | .. | 127 (15%) | 0·09 |
| Oesophagus | .. | 13 (1%) | .. | 177 (21%) | 0·69 |
| Kidney | .. | 43 (3%) | .. | 0 | 0·25 |
| Lymphoma‡ | .. | 249 (17%) | .. | 0 | 0·65 |
| Bladder or urothelial | .. | 7 (<1%) | .. | 11 (1%) | 0·09 |
| Uterus | .. | 8 (<1%) | .. | 2 (<1%) | 0·05 |
| Cervical | .. | 16 (1%) | .. | 2 (<1%) | 0·11 |
| Ovarian | .. | 5 (<1%) | .. | 33 (4%) | 0·25 |
| Other§ | .. | 57 (4%) | .. | 0 | 0·29 |
| Very high | .. | 181 (13%) | .. | 153 (18%) | 0·16 |
| Pancreas | .. | 118 (8%) | .. | 116 (14%) | 0·18 |
| Stomach | .. | 63 (4%) | .. | 37 (4%) | 0·00 |
| Newly diagnosed cancer | 1423 | 1008 (71%) | 581 | 454 (78%) | 0·17 |
| Tumour grade | 1392 | .. | 0 | .. | .. |
| 1 or 2 | .. | 868 (62%) | .. | NA | NA |
| 3 or 4 | .. | 524 (38%) | .. | NA | NA |
| Tumour stage (UICC or Ann Arbor) | 1354 | .. | 826 | .. | .. |
| I | .. | 139 (10%) | .. | 10 (1%) | 0·40 |
| II | .. | 311 (23%) | .. | 48 (6%) | 0·50 |
| III | .. | 222 (16%) | .. | 258 (31%) | 0·35 |
| IV | .. | 682 (50%) | .. | 510 (62%) | 0·23 |
| Haemoglobin (g/dL) | 1416 | 12·9 (11·6–14·0) | 820 | 13·1 (11·7–14·2) | 0·11 |
| Leucocyte count (×10³ per µL) | 1416 | 7·2 (5·7–9·4) | 820 | 7·7 (6·0–9·9) | 0·10 |
| Neutrophil count (×10³ per µL) | 1112 | 4·7 (3·5–6·3) | 0 | NA | NA |
| Platelet count (×10³ per µL) | 1416 | 248 (197–309) | 819 | 280 (224–353) | 0·37 |
| D-dimer (µg/mL) | 1423 | 0·7 (0·4–1·5) | 832 | 0·94 (0·46–2·08) | 0·10 |
| Soluble P-selectin (ng/mL) | 1410 | 40·3 (30·9–50·6) | 828 | 34·0 (26·0–43·0) | 0·40 |
| Fibrinogen (mg/dL) | 1416 | 392 (324–489) | 0 | NA | NA |
| Factor VIII activity (%) | 1417 | 183 (143–234) | 0 | NA | NA |
| Prothrombin fragment 1·2 (pM/L) | 1406 | 232 (169–330) | 0 | NA | NA |
| Peak of thrombin generation (nM) | 1415 | 392 (207–542) | 0 | NA | NA |
| Velocity index of thrombin generation (nM/min) | 1415 | 81·3 (31·8–147·4) | 0 | NA | NA |

(Table 1 continues on next page)

| | CATS (development cohort; n=1423) | | MICA (validation cohort; n=832) | | SMD* |
|---|---|---|---|---|---|
| | n† | Summary measure | n† | Summary measure | |
| (Continued from previous page) | | | | | |
| Khorana score | 1411 | 1 (0–2) | 814 | 1 (1–2) | 0·28 |
| 0 points | .. | 441 (31%) | .. | 163 (20%) | 0·26 |
| 1 point | .. | 512 (36%) | .. | 292 (36%) | 0·01 |
| 2 points | .. | 322 (23%) | .. | 242 (30%) | 0·16 |
| ≥3 points | .. | 136 (10%) | .. | 117 (14%) | 0·15 |
| Mean (SD) | 1411 | 1·1 (1·0) | 814 | 1·4 (1·0) | 0·28 |
| Vienna update to the Khorana score | 1402 | .. | 810 | .. | 0·21 |
| 0 points | .. | 317 (23%) | .. | 107 (13%) | 0·25 |
| 1 point | .. | 406 (29%) | .. | 232 (29%) | 0·01 |
| 2 points | .. | 359 (26%) | .. | 240 (30%) | 0·09 |
| 3 points | .. | 203 (14%) | .. | 145 (18%) | 0·09 |
| ≥4 points | .. | 117 (8%) | .. | 86 (11%) | 0·08 |

Data are median (IQR) or n (%), unless otherwise stated. CATS=Vienna Cancer and Thrombosis Study. MICA=Multinational Cohort Study to Identify Cancer Patients at High Risk of Venous Thromboembolism. SMD=standardised mean difference. VTE=venous thromboembolism. NA=not available. UICC=Union Internationale pour le Lutte Contre le Cancer. *An SMD of 1 indicates that the mean of the data would be one SD higher in CATS than in MICA or in MICA than in CATS; we considered SMDs of greater than 0·2 to indicate a potentially relevant difference between CATS and MICA.[18] †Data are number of patients with data available. ‡Includes only lymphoma that could be staged with Ann Arbor criteria. §Most other sites were sarcomas and testicular germ-cell tumours.

***Table 1:*** **Baseline characteristics of the study cohorts**

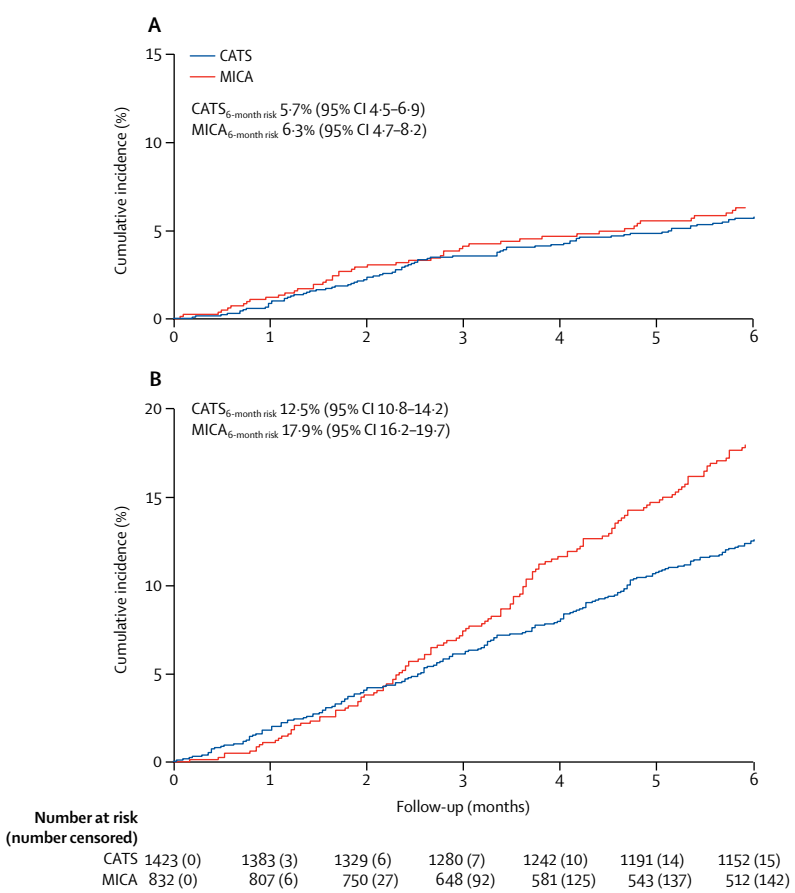


***Figure 1:*** **Cumulative incidence of venous thromboembolism (A) and mortality (B) in CATS and MICA**
CATS=Vienna Cancer and Thrombosis Study. MICA=Multinational Cohort Study to Identify Cancer Patients at High Risk of Venous Thromboembolism.

c-index of the model in MICA was 0·68 (95% CI 0·62–0·74; figure 2B), showing that the predicted risk of venous thromboembolism in MICA was in agreement with the observed incidence. At a 10% cutoff for predicted 6-month risk of venous thromboembolism, the sensitivity was 21% (95% CI 10–35), the specificity was 87% (85–90), the positive predictive value was 9% (4–16), and the negative predictive value was 95% (93–96). At a cutoff of 15%, the sensitivity of the model was 8% (2–20), the specificity was 99% (98–99), the positive predictive value was 29% (8–58), and the negative predictive value was 95% (93–96).

Of the five constituents of the Khorana score, only tumour-site category was significantly associated with risk of venous thromboembolism in both CATS and MICA (table 2). The c-indices of the Khorana score for prediction of the 6-month risk of venous thromboembolism were 0·61 (95% CI 0·51–0·70) in CATS and 0·56 (0·50–0·63) in MICA, which were lower than the corresponding c-indices for our clinical prediction model. For both CATS and MICA, the c-indices for our clinical prediction were similar to those for the Vienna modification of the Khorana score (0·66 [95% CI 0·58–0·73] in CATS and 0·63 [0·55–0·70] in MICA). Applying our model instead of the Khorana score reclassified 31% of patients in CATS correctly according to whether they did or did not develop venous thromboembolism (population-weighted NRI was 0·31).

The decision-curve analysis showed that the model had greater clinical utility for thromboprophylaxis indication than did the strategies of treat all or treat none (figure 4). Particular benefit was seen for physicians who would

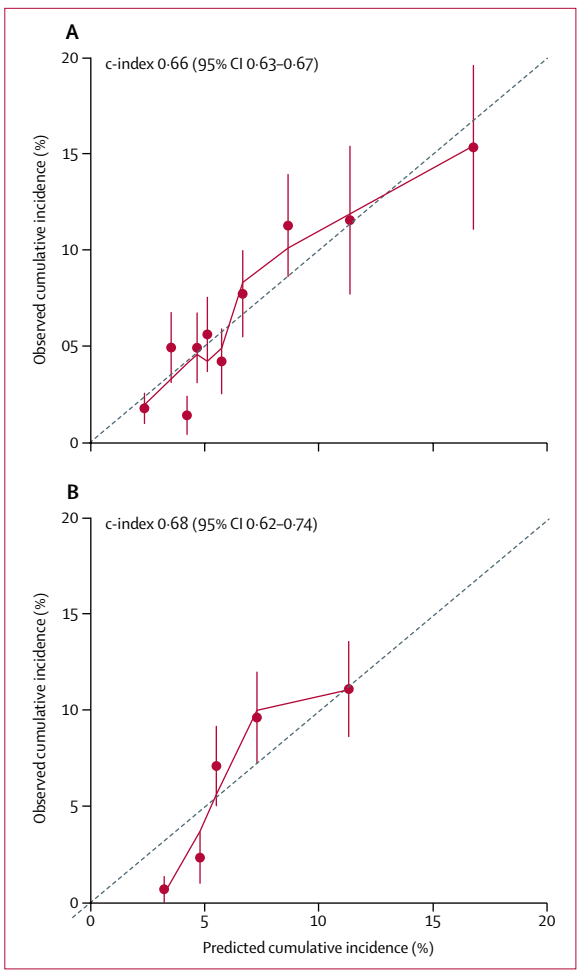


***Figure 2:*** **Cross-validated calibration plots of the clinical prediction model in CATS (A) and MICA (B)**
These graphs plot observed against predicted 6 months' venous thromboembolism risk within deciles (CATS) or quintiles (MICA) of the score's linear predictor. A smaller distance of the scatter points from the dotted line indicates better calibration. Error bars are 95% CIs. CATS=Vienna Cancer and Thrombosis Study. MICA=Multinational Cohort Study to Identify Cancer Patients at High Risk of Venous Thromboembolism.

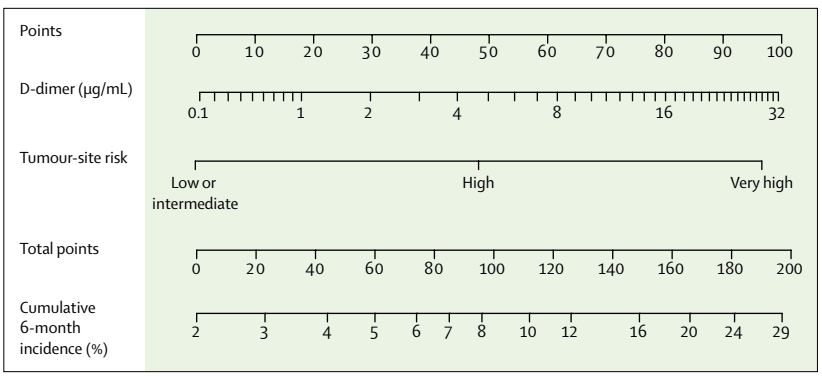


***Figure 3:*** **Nomogram for predicting the 6-month risk of venous thromboembolism**
Points for D-dimer concentration and tumour-site risk category can be obtained by calibrating with the point caliper, and then combined to obtain a total score that can be calibrated with the cumulative 6-month incidence scale. The equation for predicting 6-month risk of venous thromboembolism is provided in the appendix (p 7).

| | Multivariable SHR in CATS | Multivariable SHR in MICA |
|---|---|---|
| Tumour-site risk category | | |
| Low or intermediate | Ref (1·00) | Ref (1·00) |
| High | 1·99 (1·00–3·94; 0·05) | 2·29 (1·09–4·81; 0·028) |
| Very high | 4·54 (2·15–9·62; <0·0001) | 2·00 (0·82–4·87; 0·13) |
| BMI ≥35 kg/m² | 1·85 (0·66–5·15; 0·24) | NA* |
| Platelet count ≥350 × 10⁹/L | 1·15 (0·65–2·03; 0·63) | 1·25 (0·69–2·28; 0·46) |
| Haemoglobin level <10 g/dL or ESA use | 1·47 (0·81–2·68; 0·21) | 1·45 (0·59–3·59; 0·42) |
| White blood cell count >11 × 10⁹/L | 1·03 (0·54–1·96; 0·93) | 0·80 (0·37–1·73; 0·56) |

Data are SHR (95% CI; p value). Results were estimated with multivariable Fine and Gray competing risk regression models, considering death from any cause except fatal venous thromboembolism as the competing event of interest. SHR=subdistribution hazard ratio. CATS=Vienna Cancer and Thrombosis Study. MICA=Multinational Cohort Study to Identify Cancer Patients at High Risk of Venous Thromboembolism. BMI=body-mass index. NA=not applicable. ESA=erythropoesis-stimulating agents. *The SHR for BMI could not be estimated because no events were observed in the 26 MICA patients with a BMI of ≥35 kg/m².

***Table 2:*** **Associations between individual Khorana score items and 6-month venous thromboembolism risk in CATS and MICA**

consider thromboprophylaxis for patients with 6-month venous thromboembolism risks in the range of 6–11%.

## Discussion

We developed a novel clinical prediction model for venous thromboembolism in ambulatory patients with various types of solid cancer in a large prospective cohort, and then externally validated the model in an independent prospective cohort. Analysis of a large number of clinical and laboratory parameters produced a simple model that included only one clinical variable (tumour-site category) and one biomarker (D-dimer), which can be routinely tested for in most hospitals and cancer treatment centres worldwide. The resulting nomogram was able to discriminate between patients who did and did not develop venous thromboembolism during 6 months of follow-up, and was appropriately calibrated. Decision-curve analysis showed that use of the model to select those patients who would benefit from thromboprophylaxis would provide greater clinical utility, by reducing the risks of venous thromboembolism and bleeding events caused by unnecessary thromboprophylaxis, as compared with treat-all or treat-none approaches. The model is available as a paper-based nomogram and as an online risk calculator.

For more on the **risk calculator** see catscore.meduniwien.ac.at

This novel and simple tool might enable clinicians to identify those ambulatory patients with solid cancers who have a 6-month venous thromboembolism risk of 10–15% or more and thus might benefit from thromboprophylaxis, and those at a very low risk of venous thromboembolism, in whom the increased risk of bleeding due to thromboprophylaxis would outweigh the benefits.[2]

Several clinical prediction models for venous thromboembolism in the oncological setting exist,[24] but these do not distinguish between high-risk and low-risk patients accurately enough. The most widely used is the Khorana score, which has had external validation in several

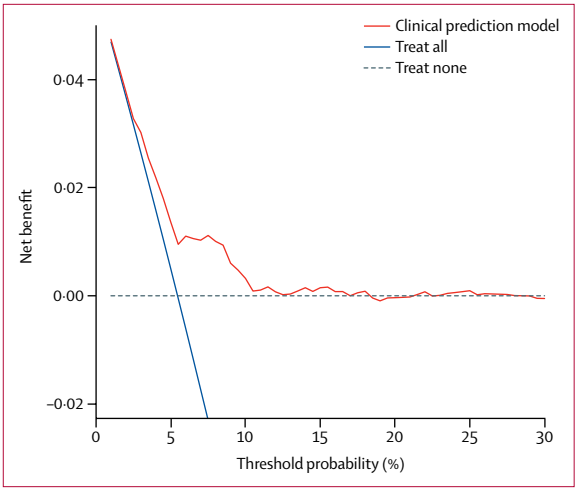


***Figure 4:*** **Decision-curve analysis for primary thromboprophylaxis in CATS**
The threshold probability represents the predicted 6-month risk of venous thromboembolism in CATS for recommending primary thromboprophylaxis. The net clinical benefit balances the risk of venous thromboembolism with the potential harms of unnecessary thromboprophylaxis and was calculated as the true-positive rate minus the weighted false-positive rate. CATS=Vienna Cancer and Thrombosis Study.

studies[5,8] and is endorsed in guidelines of the American Society of Clinical Oncology.[6] We incorporated the most important item of the Khorana score, tumour-site category,[25] into our clinical prediction model. The 6-month prediction time window chosen for our model covers the period of highest incidence of venous thromboembolism in ambulatory patients with solid cancers.[7] Unlike our previous clinical prediction model,[7] which included blood count parameters that can vary for several reasons—particularly depending on the chemotherapy regimen used—we included D-dimer concentrations in the model, which are not affected by chemotherapy.[26] We did not include soluble P-selectin concentrations, which are used in scores such as the Vienna modification of the Khorana score, in our model because this biomarker did not reach our predefined cutoff for a meaningful predictive parameter (SHR <1·25). Furthermore, methods for the measurement of soluble P-selectin are not readily available in routine laboratory settings, making widespread use of this test unfeasible in daily clinical care.

Previous studies have assessed the performance of various biomarkers to predict risk of developing cancer-associated venous thromboembolism,[12,27] including some—such as soluble vascular endothelial growth factor and thrombin—which are not routinely available outside of research environments. Using a least absolute shrinkage and selection operator approach, we identified D-dimer as the strongest prognostic biomarker for venous thromboembolism in patients with cancer of all previously tested biomarkers. D-dimer has been validated across multiple cohorts for exclusion of venous thromboembolism in diagnostic settings and as an independent venous thromboembolism risk factor in prognostic settings for patients with and without cancer.[28] This test is widely available in health-care facilities. Although the CATS and MICA studies used D-dimer assays from different manufacturers, the prognostic performance of D-dimer testing was consistent, which is reassuring for the use of other D-dimer assays with our clinical prediction model. However, other D-dimer assays need to be validated for use in our model before they can be used to predict cancer-associated venous thromboembolism.

Risk thresholds for considering prophylactic anti-coagulation to prevent venous thromboembolism in patients with cancer are subjective from the perspective of both the physician and the patient. It has previously been shown that the Khorana score can be used to select those patients at high risk of venous thromboembolism who would benefit from thromboprophylaxis, leading to a reduction in incidence of venous thromboembolism, but with increased risk of clinically relevant bleeding.[29] We assessed the clinical utility of our model using a decision-curve analysis,[30] and found that it could be useful to predict those patients with a venous thromboembolism risk greater than 5–15% who would benefit from thromboprophylaxis. However, further research is needed to assess the benefit of thromboprophylaxis in patients with cancer who are at the highest risk of developing venous thromboembolism, most appropriately in risk-adapted trials. Until results from such trials become available, our data suggest useful thresholds that might be considered for primary thromboprophylaxis.

Nonetheless, reasonable evidence exists to suggest that thromboprophylaxis halves the absolute risk of venous thromboembolism in patients with cancer.[2] On the basis of this assumed absolute risk reduction, the numbers needed to treat to prevent one cancer-associated venous thromboembolism would be 40 or more patients in the 2–5% risk range, between 20 and 40 in the 5–10% risk range, between 14 and 19 in the 10–15% risk range, and fewer than 14 in the 15% or higher risk range.[2] We posit that thromboprophylaxis is justified for patients with cancer who have a predicted 6-month risk of developing venous thromboembolism of 15% or higher, and perhaps even for those with a 10–15% risk in light of American College of Clinical Pharmacy guidelines that recommend long-term anticoagulation in patients with venous thromboembolism in the absence of cancer who have a risk of recurrence of about 10% at 12 months.[31] Furthermore, in patients without cancer who are undergoing orthopaedic surgery, even a risk of symptomatic venous thromboembolism of 5% or lower is considered relevant for initiation of thromboprophylaxis for up to 6 weeks.[32]

Our study has some limitations that undermine its generalisability. First, D-dimer assays other than those used in the CATS and MICA studies could reveal different results and thus should be validated separately.

Second, because we considered a composite of deep vein thrombosis and pulmonary embolism as the primary outcome, we cannot comment on the validity of the clinical prediction model for these outcomes separately. Third, the patients in CATS and MICA were recruited from academic centres, and so probably do not reflect the full spectrum of patients with cancer. A considerable proportion of patients with oesophageal cancer were enrolled in MICA, whereas these patients were under-represented in CATS and in the Khorana score. By contrast, all the patients with lymphoma included in our study were enrolled in CATS, and we did not include data from any patients with lymphoma enrolled in MICA. Therefore, the generalisability of the model for patients with oesophageal carcinoma or lymphoma might be limited. We did not include patients with high-grade gliomas or multiple myelomas so our model could not be used in these patient groups. Furthermore, 30% of the patients in CATS did not have newly diagnosed cancer but a long history of cancer, albeit without recent chemotherapy, and 70% of patients in MICA were enrolled after the start of chemotherapy, thus weakening the comparison with the Khorana score. However, the applicability of our model to patients who have already started chemotherapy might be considered a strength, potentially broadening its application in daily practice.

In conclusion, we developed and externally validated a novel clinical prediction model for venous thromboembolism in ambulatory patients with solid cancers. With inclusion of one clinical factor (tumour-site risk category) and one biomarker (D-dimer), our simple model was able to discriminate between patients at low and high risk of venous thromboembolism. Additionally, this model has the potential for use in selection of patients with cancer who might benefit from thromboprophylaxis.

#### Contributors

IP, GH, FP, and CA designed this study. IP and CCZ designed CATS, and NvE and HRB designed MICA. FP, JR, E-MR, and CA contributed patients to CATS, and NvE, MDN, GC-M, and NK contributed patients to MICA. FP was responsible for study coordination. GH, NvE, and FP did the statistical analysis in MICA, and GH and FP contributed to the statistical analysis in CATS. IP, NvE, and FP wrote the first draft of the manuscript. All authors critically reviewed the draft and approved the final version for publication.

#### Declaration of interests

We declare no competing interests.

#### Acknowledgments

CATS was funded by the Austrian National Bank Memorial Fund (grant numbers 10935 and 12739) and the Austrian Science Fund (special research programme FWF-SFB54 InThro). MICA was funded by unrestricted grants from the participating hospitals. We thank Ankie Kleinjan, Pieter W Kamphuisen, Hans-Martin Otten, and Isabelle Mahé for enrolment of patients in MICA.

#### References

1 Ay C, Pabinger I, Cohen AT. Cancer-associated venous thromboembolism: burden, mechanisms, and management. *Thromb Haemost* 2017; **117:** 219–30.

2 Di Nisio M, Porreca E, Candeloro M, et al. Primary prophylaxis for venous thromboembolism in ambulatory cancer patients receiving chemotherapy. *Cochrane Database Syst Rev* 2016; **12:** CD008500.

3 Prandoni P, Lensing AWA, Piccioli A, et al. Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis. *Blood* 2002; **100:** 3484–88.

4 Ay C, Pabinger I. VTE risk assessment in cancer. Who needs prophylaxis and who does not? *Hamostaseologie* 2015; **35:** 319–24.

5 Khorana AA, Kuderer NM, Culakova E, Lyman GH, Francis CW. Development and validation of a predictive model for chemotherapy-associated thrombosis. *Blood* 2008; **111:** 4902–07.

6 Lyman GH, Khorana AA, Kuderer NM, et al. Venous thromboembolism prophylaxis and treatment in patients with cancer: American Society of Clinical Oncology clinical practice guideline update. *J Clin Oncol* 2013; **31:** 2189–204.

7 Ay C, Dunkler D, Marosi C, et al. Prediction of venous thromboembolism in cancer patients. *Blood* 2010; **116:** 5377–82.

8 Verso M, Agnelli G, Barni S, Gasparini G, LaBianca R. A modified Khorana risk assessment score for venous thromboembolism in cancer patients receiving chemotherapy: the Protecht score. *Intern Emerg Med* 2012; **7:** 291–92.

9 Pelzer U, Sinn M, Stieler J, Riess H. Primary pharmacological prevention of thromboembolic events in ambulatory patients with advanced pancreatic cancer treated with chemotherapy? *Dtsch Med Wochenschr* 2013; **138:** 2084–88 (in German).

10 van Es N, Di Nisio M, Cesarman G, et al. Comparison of risk prediction scores for venous thromboembolism in cancer patients: a prospective cohort study. *Haematologica* 2017; **102:** 1494–501.

11 Gerotziafas GT, Taher A, Abdel-Razeq H, et al. A predictive score for thrombosis associated with breast, colorectal, lung, or ovarian cancer: the prospective COMPASS-Cancer-Associated Thrombosis study. *Oncologist* 2017; **22:** 1222–31.

12 Pabinger I, Thaler J, Ay C. Biomarkers for prediction of venous thromboembolism in cancer. *Blood* 2013; **122:** 2011–18.

13 Ay C, Simanek R, Vormittag R, et al. High plasma levels of soluble P-selectin are predictive of venous thromboembolism in cancer patients: results from the Vienna Cancer and Thrombosis Study (CATS). *Blood* 2008; **112:** 2703–08.

14 Ay C, Vormittag R, Dunkler D, et al. D-dimer and prothrombin fragment 1 + 2 predict venous thromboembolism in patients with cancer: results from the Vienna Cancer and Thrombosis Study. *J Clin Oncol* 2009; **27:** 4124–29.

15 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015; **13:** 1.

16 Riedl J, Preusser M, Nazari PMS, et al. Podoplanin expression in primary brain tumors induces platelet aggregation and increases risk of venous thromboembolism. *Blood* 2017; **129:** 1831–39.

17 Thaler J, Ay C, Kaider A, et al. Biomarkers predictive of venous thromboembolism in patients with newly diagnosed high-grade gliomas. *Neuro Oncol* 2014; **16:** 1645–51.

18 Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015; **34:** 3661–79.

19 Ay C, Posch F, Kaider A, Zielinski C, Pabinger I. Estimating risk of venous thromboembolism in patients with cancer in the presence of competing mortality. *J Thromb Haemost* 2015; **13:** 390–97.

20 Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw* 2011; **39:** 1–13.

21 Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997; **16:** 385–95.

22 Dunkler D, Schemper M, Heinze G. Gene selection in microarray survival studies under possibly non-proportional hazards. *Bioinforma Oxf Engl* 2010; **26:** 784–90.

23 Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology* 2014; **25:** 114–21.

24 Angelini D, Khorana AA. Risk assessment scores for cancer-associated venous thromboembolic disease. *Semin Thromb Hemost* 2017; **43:** 469–78.

25 Pabinger I, Posch F. Flamethrowers: blood cells and cancer thrombosis risk. *Hematology Am Soc Hematol Educ Program* 2014; **2014:** 410–17.

26 Reitter E-M, Kaider A, Ay C, et al. Longitudinal analysis of hemostasis biomarkers in cancer patients during antitumor treatment. *J Thromb Haemost* 2016; **14:** 294–305.

27 Posch F, Thaler J, Zlabinger G-J, et al. Soluble vascular endothelial growth factor (sVEGF) and the risk of venous thromboembolism in patients with cancer: results from the Vienna Cancer and Thrombosis Study (CATS). *Clin Cancer Res* 2015; **22:** 200–06.

28 Ay C, Dunkler D, Pirker R, et al. High D-dimer levels are associated with poor prognosis in cancer patients. *Haematologica* 2012; **97:** 1158–64.

29 Khorana AA, Francis CW, Kuderer NM, et al. Dalteparin thromboprophylaxis in cancer patients at high risk for venous thromboembolism: a randomized trial. *Thromb Res* 2017; **151:** 89–95.

30 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26:** 565–74.

31 Kearon C, Akl EA, Comerota AJ, et al. Antithrombotic therapy for VTE disease: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2012; **141** (suppl 2)**:** e419S–94S.

32 Falck-Ytter Y, Francis CW, Johanson NA, et al. Prevention of VTE in orthopedic surgery patients: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2012; **141** (suppl 2)**:** e278S–325S.