Clinical and Translational Oncology (2025) 27:4031–4039
https://doi.org/10.1007/s12094-025-03890-5

RESEARCH ARTICLE

External validation of a prediction model for bleeding events in anticoagulated cancer patients with venous thromboembolism (PredictAI)

María Carmen Viñuela-Benéitez[1] · Claudia Iglesias Pérez[2] · Laura Ortega Morán[3] · Ignacio García Escobar[4] · Diego Cacho Lavín[5] · Rut Porta i Balanyà[6] · Silvia García Adrián[7] · Marta Carmona Campos[8] · Gretel Benítez López[9] · José Antonio Santiago Crespo[10] · Miriam Lobo de Mena[11] · Javier Pérez Altozano[12] · Enrique Gallardo Díaz[13] · Julia Tejerina Peces[14] · Pilar Ochoa Rivas[15] · María José Ortiz Morales[16] · Victoria Eugenia Castellón Rubio[17] · Carmen Díez Pedroche[18] · María Rosales Sueiro[19] · Felipe Gonçalves[20] · Manuel Sánchez-Cánovas[21] · Miguel Ángel Ruiz[22] · José Muñoz-Langa[23] · Pedro Pérez Segura[14] · Eva Martínez de Castro[5] · Alberto Carmona-Bayonas[21] · Paula Jiménez-Fonseca[2] · Andrés Jesús Muñoz Martín[3]

Received: 11 January 2025 / Accepted: 25 February 2025 / Published online: 26 April 2025

Abstract

Objective The objective of this study was to validate the PredictAI models for predicting major bleeding (MB) in patients with active cancer and venous thromboembolism (VTE) with anticoagulant (ACO) therapy, within 6 months after primary VTE, using an independent cohort of patients from the TESEO database.

Methods This study conducted an external validation of the PredictAI models using the international, prospective TESEO registry from July 2018 until October 2021. Data from 40 Spanish and Portuguese hospitals recruiting consecutive cases of cancer-associated thrombosis under anticoagulant treatment and without missing values regarding the model outcome or predictors were used. Patients with baseline MB or unknown MB status during follow-up were excluded for the validation analysis. Logistic regression (LR), decision tree (DT), and random forest (RF) approaches were used to validate the models.

Results Included patients from the TESEO cohort (2179 patients) had similar key demographics and clinical characteristics to the PredictAI cohort (21,227 patients). During the 6-month follow-up period, 10.9% ($n=2314$) and 5.9% ($n=129$) of patients experienced at least one MB event in the PredictAI and TESEO cohorts, respectively. Hemoglobin, metastasis, age, platelets, leukocytes, and serum creatinine were described as predictors for MB in PredictAI; the external validation results in TESEO showed statistical significance by LR and RF approaches, with ROC-AUC values of 0.59 and 0.56, respectively (both $p<0.05$).

Conclusion PredictAI models for predicting MB in anticoagulant-treated cancer patients within the first 6 months following VTE diagnosis have been externally validated. These models may be considered as a tool to guide objective decisions regarding the indication or extension of anticoagulant therapy in this population.

Keywords Cancer · Venous thromboembolism · Bleeding · Anticoagulation · Natural language processing

Introduction

Cancer patients are at high risk for developing venous thromboembolism (VTE), which in turn is a leading cause of bleeding complications and morbimortality in this population [1–4]. From a clinical standpoint, the management of bleeding events, including major bleeding (MB), in individuals with active cancer and VTE is complex and often entails delay and stop of anticancer therapy [5, 6]. Determining the individual features that prelude the occurrence of bleeding events is crucial to identify high-risk patients and support further decisions regarding the initiation or duration of anticoagulant therapy.

Existing models and risk scores for the prediction of bleeding events in patients with VTE have shown important methodological limitations and suboptimal predictive accuracy [7, 8]. While previous studies have assessed the

Extended author information available on the last page of the article

impact of bleeding in anticoagulated cancer patients with VTE [9–14], risk factors of bleeding in this population remain largely unexplored. Although active cancer is a predictive factor of bleeding in the available models developed in patients with VTE, no specific risk scores had been developed in anticoagulant-treated cancer patients until the recent development of CAT-BLEED and B-CAT score [8, 9]. However, as with some of the aforementioned models in VTE, CAT-BLEED was developed using only clinical trial data with restrictive inclusion criteria and patients undergoing treatment with specific anticoagulant drugs. Then its generalizability and clinical utility may, thus, be largely compromised. Neither model has yet been externally validated.

To circumvent some of these limitations, Muñoz and colleagues have recently developed and published the results of the PredictAI study, which includes a machine learning (ML)-based algorithm to predict MB in anticoagulant-treated cancer patients within the first 6 months following VTE diagnosis [10]. From a target population of 21,227 patients in Spain, all potential predictors and the outcome (MB) were retrospectively extracted from the unstructured information in electronic health records (EHRs) between 2014 and 2018, using EHRead® technology (Medsavana, Madrid, Spain) which includes natural language processing (NLP) and machine learning (ML). Several statistical approaches were evaluated for the development of the predictive model, namely logistic regression (LR), decision tree classifiers (DT), and random forest (RF) models. The models identified the following predictors of MB at baseline: hemoglobin levels, presence of metastasis, patient's age, platelet count, leukocyte count, and serum creatinine. The LR model had an area under the receiver operating characteristic curve (AUC-ROC) (95% CI) of 0.60 (0.55, 0.65), the DT of 0.60 (0.55, 0.65), and the RF of 0.61 (0.56, 0.66).

The overarching goal of the present study was to externally validate the PredictAI models using an independent cohort of patients from the TESEO registry [11], managed by the Spanish Society of Medical Oncology (SEOM).

Methods

Study design and data source

To externally validate the PredictAI models, we used an independent cohort of patients from the TESEO registry which is an international, prospective registry with data from 40 Spanish and Portuguese hospitals that recruit consecutive cases of cancer-associated thrombosis since July, 2018 [11]. To optimize reliability and data completeness, the registry is managed via a web-supported platform that allows to perform targeted queries and data filtering [12]. In this regard, the study used data from TESEO, which was extracted from July 2018 to October 2021. Further details are shown in Table 1, where a comparison of data source and study design between PredictAI and TESEO is provided.

Participants

For this study, patients prospectively and consecutively included in the TESEO registry were analyzed. Eligibility criteria include: over 18 years of age; histologically confirmed malignant tumor; VTE episode, symptomatic or incidental, confirmed with an imaging technique, in the month prior or any time until 2 years after the cancer diagnosis; and sign the informed consent form [13]. Table 1 shows additional inclusion and exclusion criteria in the PredictAI and TESEO databases.

Outcome, predictors, and other clinical data

The event to predict (i.e., outcome) was MB within 6 months following VTE diagnosis, which included all events originally coded by medical oncologists in the TESEO registry as either MB or clinically relevant bleeding (CRB). MB was defined by the International Society on Thrombosis and Haemostasis (ISTH) criteria in both cohorts. In TESEO registry, the MB was collected as MB or CRB. The analyzed predictors for MB at baseline comprised those originally described in PredictAI, namely patient's age, presence of metastasis, hemoglobin levels, platelet count, leukocyte count, and serum creatinine levels [10]. All variables were directly captured from the TESEO database; the calculation of leukocyte count was inferred by adding neutrophils and lymphocytes. The assessment of both the outcome and predictors was based on routine clinical practice. The assessment of MB was blind from the influence of the predictors or the main goal of the present study.

Predictive model validation

We followed the same strategy as in PredictAI for selecting patient population to validate PredictAI models with the TESEO dataset. Then participants who had experienced a MB event at baseline, as well as patients with follow-up periods shorter than 6 months after VTE diagnosis and with no record of an MB during that period, were excluded for model development. Since the models were developed using classification algorithms, this group of patients with uncertain outcomes was excluded from the PredictAI study as well as from this external validation, as they could not be assigned to either the event or non-event groups. Patients with at least one missing value regarding the model outcome (MB) or any of the considered predictors were also excluded from further analyses. No imputation methods for missing data were performed.

Table 1 Data source, study design, and key study characteristics of both databases, for the development and validation of the predictive model

	Development database PredictAI *N* = 21,227	Validation database TESEO *N* = 2179
Data source and design		
Data source and extraction	Unstructured data in EHRs extracted using EHRead® (ML and NLP technology)	Prospective registry with structured data manually collected from EHRs
Data collection period	January 2014–December 2018	July 2018–October 2021
Study design	Retrospective data collection from patients prospectively attending the participating hospitals	Prospective data collection
Setting	Multicenter; 9 Spanish hospitals; all available departments and services	Multicenter; 40 Spanish hospitals; oncology departments
Participants	Anticoagulant-treated cancer patients with VTE	Anticoagulant-treated cancer patients with VTE
Key inclusion criteria*	- Age ≥ 18 years - VTE ± 6 months from cancer diagnosis - Ongoing treatment with anticoagulant therapy	- Age ≥ 18 years - VTE −1 month or any time after the cancer diagnosis up to 2 years
Key exclusion criteria	- Anticoagulation treatment for an indication other than VTE - Malignancy + 6 months from VTE - Non-metastatic patients with cancer diagnosis > 6 months prior to VTE - VTE while having cancer on complete remission for > 2 years - Acute leukemia	- Clinical diagnosis of TEE without radiological confirmation - Patients with no MB information < 6 months follow-up - Incomplete information on any study variables
Outcome	Major bleeding	Major or CRB
Key study characteristics		
Follow-up (months)		
Median (Q1, Q3)	8 (1, 24)	6 (2, 14)
Demographics		
Male sex *n* (%)	11,431 (53.9)	1130 (51.9)
Outcome		
Frequency of MB (%)	10.9% at 6 months	5.9% at 6 months
Predictors at baseline		
Age at inclusion		
Mean (SD)	69 (15)	65 (11)
Median (Q1, Q3)	70 (59, 80)	65 (58, 73)
Metastasis (%)	9466 (44.6)	1549 (71.1)
Hemoglobin (g/dl)		
N (%)	18,884 (89)	2179 (100)
Mean (SD)	11.8 (2.5)	11.8(1.9)
Median (Q1, Q3)	11.8 (10.2, 13.4)	11.8 (10.4, 13)
Platelet count (10^3/mm^3)		
N (%)	17,110 (80.6)	2179 (100)
Mean (SD)	244 (128)	242 (120)
Median (Q1, Q3)	224 (165, 296)	220 (160, 303)
Leukocyte count (10^3/mm^3)		
N (%)	16,120 (75.9)	2179 (100)
Mean (SD)	13.9 (51.6)	7.4 (4.9)
Median (Q1, Q3)	8.4 (6.1, 11.6)	6.4 (4.5, 9.0)
Serum creatinine (mg/dl)		
N (%)	17,071 (80.4)	2179 (100)
Mean (SD)	1.26 (2.56)	0.83 (0.32)
Median (Q1, Q3)	0.88 (0.69, 1.2)	0.78 (0.63, 0.94)

CRB clinically relevant bleeding, *MB* major bleeding, *SD* standard deviation, *TEE* thromboembolic event, *VTE* venous thromboembolism

*Both incidental and symptomatic VTE were considered in both databases as inclusion criteria

The three predictive models developed in PredictAI (i.e., DT, RF classifier, and LR) were tested for validation using the TESEO database. Model performance was assessed in terms of accuracy, precision, recall, F1, and ROC-AUC. ROC confidence intervals and *p* values were computed using the approach proposed by Green & Sweets [14], retaken by Hanley & McNeil [15]. ROC curves were compared using the statistic proposed by Hanley & McNeil [15, 16].

Data analysis

Descriptive tables were elaborated to show the distribution of variables in each set. Categorical variables were presented as frequencies and numerical variables as median and interquartile range (Q1, Q3). Predictive model metrics were presented in tables. ROC curves were displayed as plots. Data were analyzed and represented using IBM SPSS (version 26) and Python (version 3.7.12) [17].

Sample size

A priori sample size calculations were not performed for the validation of the predictive model. In PredictAI, it was estimated that the minimum sample size required to establish the risk of MB in the target population was 1050 patients. This was estimated considering the development of a multiple LR model with an observed small effect size (0.05), a significance level of 0.05, and a statistical power of 0.9 [10].

Ethical considerations

All methods and analyses followed legal and regulatory requirements and generally accepted research practices described in the Helsinki Declaration in its latest edition. Data were collected from the TESEO registry of patients with cancer-related thrombosis [11, 12]. The development of TESEO was originally approved by a multicenter Research Ethics Committee of all the Autonomous Communities and participating centers; the study was classified by the Spanish Agency of Medicines and Medical Devices as a post-marketing, prospective follow-up study [12]. Consent forms were signed by all patients who were still alive at the time of data collection. Results here are reported following the Transparent Reporting of a multivariate prediction model for Individual Prognosis or Diagnosis based on artificial intelligence (TRIPOD + AI) guidelines [18].

Results

PredictAI and TESEO patients' characteristics

The TESEO cohort used to conduct the external validation comprised 2,179 patients with active cancer and VTE under anticoagulant treatment (51.9% male, mean age = 64.7 ± 11.3 years). A direct comparison of key demographics, clinical characteristics, MB rates, and predictive variables in the PredictAI and TESEO cohorts is shown in Table 1. Additional information regarding disease-specific clinical characteristics and comorbidities at baseline are included in supplementary Table S1 and Table S2, respectively. In TESEO, the median (Q1, Q3) follow-up duration across patients was 6 (2, 14) months. During the 6-month follow-up period, at least one MB event was registered in 5.9% ($n = 129$) of patients. Among this population, 1,863 patients were finally included in the external cohort to test the PredictAI models after applying the corresponding filters.

Validation of the predictive models

When applying the predictive models developed in PredictAI to the TESEO database, the results showed a performance significantly above chance for the LR (ROC-AUC = 0.59; 95% CI = 0.53, 0.65; $p = 0.002$) and RF (ROC-AUC = 0.56; 95% CI = 0.51, 0.62; $p = 0.023$) approaches, while it yielded a ROC-AUC of 0.53 for the DT model (0.48, 0.59) (Fig. 1). The rest of the performance metrics, as well as a comparison with the internal validation metrics, are shown in Table 2.

Discussion

The present study aimed to validate the PredictAI models for predicting MB events in anticoagulant-treated cancer patients within the 6 months following VTE diagnosis [10]. This external and independent validation was conducted with the TESEO prospective registry of patients with cancer-related thrombosis. Our results support the independent validation of the original predictive models for both the LR and RF approaches.

The clinical importance of the present validation is better understood considering the lack of reliable algorithms or risk scores for the prediction of bleeding in anticoagulated patients with VTE [7]. Notably, existing models in general population such as HASBLED [19], RIETE [20], Hokusai [21], EINSTEIN [22], ACCP [23], and BACS [24] have important limitations that impact their generalizability [25]. For instance, the RIETE score only predicts bleeding

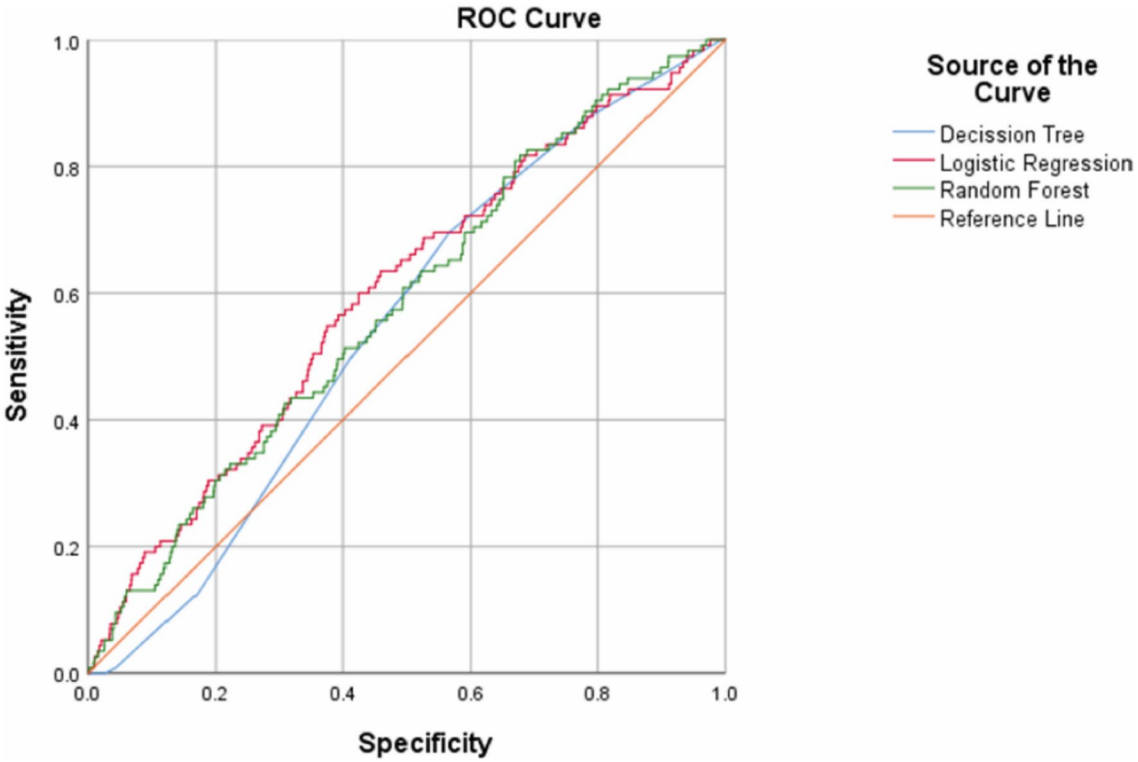

Fig. 1 Receiver operating characteristic (ROC) curves for the predictive models for MB developed in PredictAI when applied to the TESEO database. ROC curves for the decision tree model (ROC-AUC = 0.53) (blue line), logistic regression model (ROC-AUC = 0.59) (red line), and random forest model (ROC-AUC = 0.56) (green line). The yellow line represents a random classifier

Table 2 Performance metrics in the external validation of the PredictAI models

		Original validation (PredictAI)*	External validation (TESEO)
Decision tree (DT)	ROC-AUC	0.60 (0.55, 0.65)	0.53 (0.48, 0.59)
	p value	**< 0.0001**	**0.241**
	Recall	0.67	0.69
	Precision	0.57	0.069
	Accuracy	0.59	0.43
	F1 score	0.61	0.13
Logistic regression (LR)	ROC-AUC	0.60 (0.55, 0.65)	0.59 (0.53, 0.65)
	p value	**< 0.0001**	**0.002****
	Recall	0.64	0.80
	Precision	0.56	0.07
	Accuracy	0.58	0.35
	F1 score	0.60	0.12
Random forest (RF)	ROC-AUC	0.61 (0.56, 0.66)	0.56 (0.51, 0.62)
	p value	**< 0.0001**	**0.023****
	Recall	0.62	0.69
	Precision	0.57	0.07
	Accuracy	0.59	0.39
	F1 score	0.60	0.12

ROC curves are expressed including the 95% confidence interval

ROC-AUC receiver operating characteristic curve-area under the curve

*Metrics obtained in the internal validation of the model (i.e., using the validation set)**$p < 0.05$

during the 3 months following anticoagulation treatment onset [22]. The Hokusai score is only validated in patients using edoxaban, and its predictive power decreases after the first 3 months post-treatment [21]. The EINSTEIN model has limited applicability in clinical settings since it can only be calculated through a spreadsheet [22]. The ACPP incorporates vaguely defined predicting variables such as frequent falls or 'comorbidity [23]. Lastly, the BACS was developed in the subset of VTE patients with pulmonary embolism [24]. In this context, the further developed VTE-BLEED [26] aimed to solve some of these issues, and although not prospectively validated, it has already shown good performance and clinical utility [27]. In a post hoc analysis of the randomized phase III clinical trial HOKUSAI VTE Cancer, different bleeding risk assessment models were evaluated [8]. All risk scores performed poorly for bleeding (a composite of MB and CRNMB), and a pragmatic classification based only on cancer type provided a better estimate of clinically relevant bleeding risk. Based on these results, it is suggested that bleeding risk scores developed in the general population should not be used in cancer patients due to insufficient discrimination.

While active cancer is a known chief predictor of bleeding in patients with VTE, there is a lack of models that have been specifically developed and externally validated in the subpopulation of patients with cancer and VTE. Two recent scores have been developed to assess the risk of bleeding in patients diagnosed with cancer-associated thrombosis who are receiving anticoagulant therapy. De Winter and colleagues developed CAT-BLEED model for bleeding risk at 6 months following thrombosis diagnosis identified age (i.e., ≥ 75 years), creatinine clearance, and four cancer-related variables (locally advanced or metastatic stage, genitourinary cancer, recent use of anticancer treatment associated with gastrointestinal toxicity, and gastrointestinal cancer plus treatment with edoxaban) as main risk bleeding factors. Though this model is promising and represents the first bleeding risk model specifically developed in cancer-related thrombosis, it should be noted on data from the HOKUSAI VTE Cancer trial. This fact may explain the high specificity of the risk factors identified (treatment with edoxaban) which already may impact its generalizability. Moreover, it has been recently externally tested with PredictAI showing a poor performance with values of 0.53 (0.48, 0.59) [10]. The second model developed specifically for cancer-associated VTE is the B-CAT model [9]. The strength of this model, in contrast to the CAT-BLEED model, is that it has been developed in a real-world population outside of a clinical trial. However, this model has several limitations. These include the exclusion of patients with a history of VTE, the lack of a definition of minor trauma, and the non-inclusion of analytical parameters strongly associated with bleeding risk, such as thrombopenia. Another controversial aspect of this model is the definition of the target bleed in the study. In addition to MB, the authors only consider CRNMB requiring hospitalization, excluding CRNMB that result in non-admission.

In PredictAI, the inclusion of a larger and more representative series of patients and the use of RWD extracted using artificial intelligence techniques to build the model aimed to bypass these constraints. In the original model, the best performing algorithm based on the RF approach identified three of the variables featured in CAT-BLEED (i.e., patient's age, presence of metastasis, and creatinine levels) plus three additional laboratory values that are routinely collected in all patients in this clinical population, namely hemoglobin levels, platelet count, and leukocyte count. Interestingly, PredictAI models outperformed the CAT-BLEED score. Moreover, CAT-BLEED did not perform above chance when evaluated with PredictAI data [10].

The performance metrics obtained here in the external validation of PredictAI showed high recall values yet low precision. From a clinical standpoint, however, the priority of this model should be to detect most patients at risk for bleeding (i.e., high recall) to plan early prevention and surveillance, even if many of them will not eventually experience MB events. PredictAI obtained an AUC-ROC (95% CI) for the LR of 0.60 (0.55, 0.65), 0.60 (0.55, 0.65) for the DT and 0.61 (0.56, 0.66) for the RF. A minor decrease in these metrics is observed in the present validation, and this could be accounted for by methodological differences as well as patients' characteristics in the PredictAI vs. TESEO databases. From a methodological standpoint, it should be noted that while the study variables in PredictAI were obtained from patients' EHRs in all hospital departments and services available using ML and NLP, TESEO consists of a prospective registry of manually collected data in oncology departments. Thus, the definition and collection of the model's outcome and predictive variables may not be homogeneous across databases. These methodological differences may also explain the smaller rates of MB found in TESEO vs. PredictAI, where bleeding events not documented as MB in patients' EHRs may have been classified as such by the NLP system (e.g., incidental findings on imaging reports regarding tumoral bleeding) [10]. Key predictive variables in the original model such as demographics and laboratory parameters were similar in the two cohorts. However, the percentage of patients with metastasis in TESEO was markedly larger than in PredictAI, most likely due to differences in the data source and study setting; because patients in TESEO were recruited through oncology departments, the disease status in these patients is expected to be more severe, including larger prevalence of metastasis. Despite the aforementioned differences between the two studies and populations, the models still performed above chance for the detection of MB events in the independent patient cohort.

Although the PredictAI models performed significantly above chance using the independent TESEO cohort and represent an important advance from previously bleeding risk models in anticoagulated cancer patients with VTE, their application in clinical settings needs to be facilitated through the development of a clinical score or application to evaluate the risk of bleeding at the individual patient level. Moreover, prior to their application in routine clinical practice, the current models would benefit from further validations in other patient cohorts. Additional model training using more data, variables, and multilayer databases may allow for the refinement of the models to obtain more robust metrics and reduce the number of false positives.

Our results support the independent validation of the models developed in PredictAI, which predict MB within 6 months after VTE diagnosis in anticoagulated cancer patients. Given the lack of reliable predictive models for bleeding in patients with cancer and VTE, these results hold great promise for guiding objective decisions regarding the extension of anticoagulant therapy in this population. Future steps may involve further validation of the present model in other patient cohorts and the generation of a clinical tool to facilitate its use in healthcare settings.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12094-025-03890-5.

Acknowledgements The authors thank the collaboration of the subinvestigators in the different participating hospitals. We thank all members in the Savana Research Group for their methodological and technical support: Miren Taberna, Víctor Fanjul, Diego Benavent, Judith Marin-Corral, David Casadevall, Sebastian Menke, Ignacio Salcedo, Claudia Maté, and Natalia Polo. Medical writing support was provided by Carlos Del Rio-Bermudez.

Author contributions MCV and AJMM conceived and designed the study. MCV and AM performed the data collection and wrote the manuscript. MR performed the statistical analysis. All authors interpreted data, drafted, reviewed, edited, and approved the manuscript.

Funding This study was funded by unrestricted grant from BMS-Pfizer Alliance to Cancer & Thrombosis Section of the Spanish Society of Medical Oncology (SEOM).

Data availability Data are available on reasonable request to the authors. Thereafter, the committee of the project, together with the Ethics Committee of the hospitals involved, will assess the proposal and potentially proceed to the data sharing.

Code availability All code for analysis is available upon reasonable request from the corresponding author.

Declarations

Conflict of interests Personal financial interests: *ACB* reports receiving lecture grants from Esteve, Lilly, and Astellas; travel support from Amgen; and research funding from HRA Pharma-Esteve. *AJMM*—consulting or advisory role: Regeneron, Sanofi, Bristol-Myers Squibb/Pfizer, Leo Pharma, Daiichi Sankyo, Incyte, AstraZeneca, MSD Oncology, Lilly, Celgene, Novocure, Roche, Servier; Speakers' Bureau: Rovi, STADA, Menarini, Bayer; research funding: Celgene, Sanofi, Leo Pharma and Rovi; patents: risk assessment model in venous thromboembolism in cancer patients; travel, accommodation and expenses: AstraZeneca, Amgen, Merck Serono, Roche, Celgene, Pfizer. *DCL*—consultant or advisory role: AstraZeneca, BMS; speaking: AstraZeneca, BMS, Leo Pharma, Rovi. *EGD*—consultant or advisory role: Sanofi, Janssen, Astellas, Bayer, Ipsen, Pfizer, Roche, Novartis, Eisai, Recordati, BMS, MSD, AstraZeneca, Merck, Rovi, Daiichi Sankyo, Techdow, Lilly, GSK, advanced accelerator applications; speaking: Astellas, Janssen, Sanofi, Bayer, Ipsen, Pfizer, Roche, BMS, Rovi, Daiichi Sankyo, Leo Pharma, Eisai, Boehringer Ingelheim, Merck, EUSA Pharma/Recordati, Lilly, advanced accelerator applications, GSK; grant support: Astellas, Janssen, Sanofi, Bayer, Ipsen, Ferrer, Pfizer, Roche, GSK, BMS, MSD, Merck, Recordati, AstraZeneca, Advanced accelerator applications; travel/accommodation expenses: Astellas, Janssen, Sanofi, BMS, Bayer, Ipsen, Roche, Novartis, Pfizer, Eisai, AstraZeneca, Leo Pharma. *GBL*—speaking: MSD, SUNpharma, Pfizer, Pierre Fabre, Bristol Mayer, Roche, Merck, AstraZeneca. IGE—consultant: AstraZeneca, Amgen, Merck, Leo Pharma, Rovi. *JML*—speaking: GSK, MSD, AstraZeneca; travel, accommodations: Servier, MSD. *LOM*—speaking: Leo Pharma and Rovi. *MLM*—honoraria: Servier, Leo Pharma; travel, accommodations and expenses: AstraZeneca, Novartis, Ipsen, Servier, Pharma Mar, Roche. *MAR*—MR's department has received a grant from SEOM (Spanish Society of Medical Oncology) for the assessment and completion of statistical analysis. *MSC*—consultant or advisory role: KyowaKirin; research funding: Leo Pharma; speaking: Leo Pharma. Sanofi. Lundbeck; other: financial support for attending symposia (Sanofi, MSD, Esteve, Amgen, Servier, Angelini, Leo Pharma), financial support for educational programs (Angelini, Sanofi, Rovi, Leo Pharma, Servier, Merck), remunerations for authorship in clinical cases journals/book chapters/complete books (KyowaKirin, Mylan, Leo Pharma). *PJF*—honoraria for speakers' bureau participation, and serving on advisory boards from Astellas, AstraZeneca, Bristol-Myers Squibb (BMS), Esteve, Merck Sharp & Dohme (MSD), Novartis, Nutricia, Pfizer, Rovi, Takeda, and Viatri and research grants from Astellas, AstraZeneca, BMS, and MSD. *PPS*—consultant or advisory role: Roche, BMS, MSD, AstraZeneca, Merck, Rovi, Novocure, LeoPharma; speaking: Roche, BMS, MSD, AstraZeneca, Merck, Rovi, Novocure, LeoPharma; grant support: Roche, BMS, MSD, AstraZeneca, Merck, Rovi, Novocure, LeoPharma; travel/accommodation expenses: Roche, BMS, MSD, AstraZeneca, Merck, Rovi, Novocure, LeoPharma. *CIP, CDP, DB, EMC, IGE, JPA, JASC, JTP, MCC, MCVB, MJOM, MRS, POR, RPB, SGA, VECR*: the rest of authors declare no competing interests for this manuscript.

Human and animal rights All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent All participants provided informed consent prior to their participation.

References

1. Khorana AA, Francis CW, Culakova E, Kuderer NM, Lyman GH. Thromboembolism is a leading cause of death in cancer patients receiving outpatient chemotherapy. J Thromb Haemost. 2007;5(3):632–4. https://doi.org/10.1111/j.1538-7836.2007.02374.x.
2. Khorana AA. Venous thromboembolism and prognosis in cancer. Thromb Res. 2010;125(6):490–3. https://doi.org/10.1016/j.thromres.2009.12.023.
3. Elting LS, Escalante CP, Cooksley C, Avritscher EB, Kurtin D, Hamblin L, et al. Outcomes and cost of deep venous thrombosis among patients with cancer. Arch Intern Med. 2004;164(15):1653–61. https://doi.org/10.1001/archinte.164.15.1653. (**Erratum in: Arch Intern Med. 2008 Jun 9;168(11):1218**).
4. Prandoni P, Lensing AW, Piccioli A, Bernardi E, Simioni P, Girolami B, et al. Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis. Blood. 2002;100(10):3484–8. https://doi.org/10.1182/blood-2002-01-0108.
5. Patel HK, Khorana AA. Anticoagulation in cancer patients: a summary of pitfalls to avoid. Curr Oncol Rep. 2019;21(2):18. https://doi.org/10.1007/s11912-019-0767-5.
6. Lyman GH, Bohlke K, Khorana AA, Kuderer NM, Lee AY, Arcelus JI, American Society of Clinical Oncology, et al. Venous thromboembolism prophylaxis and treatment in patients with cancer: American society of clinical oncology clinical practice guideline update 2014. J Clin Oncol. 2015;33(6):654–6. https://doi.org/10.1200/JCO.2014.59.7351.
7. de Winter MA, van Es N, Büller HR, Visseren FLJ, Nijkeuter M. Prediction models for recurrence and bleeding in patients with venous thromboembolism: a systematic review and critical appraisal. Thromb Res. 2021;199:85–96. https://doi.org/10.1016/j.thromres.2020.12.031.
8. de Winter MA, Dorresteijn JAN, Ageno W, Ay C, Beyer-Westendorf J, Coppens M, et al. Estimating bleeding risk in patients with cancer-associated thrombosis: evaluation of existing risk scores and development of a new risk score. Thromb Haemost. 2022;122(5):818–29. https://doi.org/10.1055/s-0041-1735251.
9. Cohen AT, Wallenhorst C, Choudhuri S, Nassar A, Pollock KG, Martinez C. A novel risk prediction score for clinically significant bleeding in patients anticoagulated for venous thromboembolism with active cancer. Thromb Haemost. 2024;124(4):324–36. https://doi.org/10.1055/a-2145-7238.
10. Muñoz Martín AJ, Lecumberri R, Souto JC, Obispo B, Sanchez A, Aparicio J, et al. Prediction model for major bleeding in anticoagulated patients with cancer-associated venous thromboembolism using machine learning and natural language processing. Clin Transl Oncol. 2024. https://doi.org/10.1007/s12094-024-03586-2. (**Epub ahead of print. Erratum in: Clin Transl Oncol. 2024 Nov 19**).
11. Muñoz Martín AJ, Jiménez-Fonseca P, Carmona-Bayonas A, Martínez de Castro E, Muñoz Langa J, Pérez Segura P, et al. Cancer and thrombosis section, Spanish society of medical oncology (SEOM). TESEO, cancer-associated thrombosis registry from the Spanish society of medical oncology (SEOM). Clin Transl Oncol. 2020;22(8):1423–4. https://doi.org/10.1007/s12094-019-02265-x.
12. Carmona-Bayonas A, Gómez D, Martínez de Castro E, Pérez Segura P, Muñoz Langa J, Jimenez-Fonseca P, et al. A snapshot of cancer-associated thromboembolic disease in 2018–2019: first data from the TESEO prospective registry. Eur J Intern Med. 2020;78:41–9. https://doi.org/10.1016/j.ejim.2020.05.031.
13. Muñoz-Langa J, Jimenez-Fonseca P, Carmona-Bayonas A, de Castro EM, Pérez-Segura P, Cánovas MS, et al. Rationale, design and methodology of TESEO study: a registry of thrombosis and neoplasia of SEOM (Spanish society of medical oncology). Clin Transl Oncol. 2021;23(4):799–811. https://doi.org/10.1007/s12094-020-02472-x. (**Epub 2020 Aug 13**).
14. Green DM, Swets JA. Signal detection theory and psychophysics. New York: Wiley; 1966.
15. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36. https://doi.org/10.1148/radiology.143.1.7063747.
16. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148(3):839–43. https://doi.org/10.1148/radiology.148.3.6878708.
17. IBM Corp. Released. IBM SPSS statistics for windows, version 26.0. Armonk: IBM Corp; 2019.
18. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55–63. https://doi.org/10.7326/M14-0697. (**Erratum. in: Ann Intern Med. 2015 Apr 21;162(8):600.doi:10.7326/L15-0078-4. PMID: 25560714**).
19. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro heart survey. Chest. 2010;138(5):1093–100. https://doi.org/10.1378/chest.10-0134.
20. Ruíz-Giménez N, Suárez C, González R, Nieto JA, Todolí JA, Samperiz AL, RIETE Investigators, et al. Predictive variables for major bleeding events in patients presenting with documented acute venous thromboembolism. Findings from the RIETE registry. Thromb Haemost. 2008;100(1):26–31. https://doi.org/10.1160/TH08-03-0193.
21. Di Nisio M, Raskob G, Büller HR, Grosso MA, Zhang G, Winters SM, et al. Prediction of major and clinically relevant bleeding in patients with VTE treated with edoxaban or vitamin K antagonists. Thromb Haemost. 2017;117(4):784–93. https://doi.org/10.1160/TH16-11-0830. (**Epub 2017 Feb 2**).
22. Di Nisio M, Ageno W, Rutjes AW, Pap AF, Büller HR. Risk of major bleeding in patients with venous thromboembolism treated with rivaroxaban or with heparin and vitamin K antagonists. Thromb Haemost. 2016;115(2):424–32. https://doi.org/10.1160/TH15-06-0474. (**Epub 2015 Oct 29**).
23. Ageno W, Gallus AS, Wittkowsky A, Crowther M, Hylek EM, Palareti G. Oral anticoagulant therapy: antithrombotic therapy and prevention of thrombosis, 9th ed: American college of chest physicians evidence-based clinical practice guidelines. Chest. 2012;141(2 Suppl):e44S-e88S. https://doi.org/10.1378/chest.11-2292.
24. Jara-Palomares L, Jiménez D, Bikdeli B, Muriel A, Rali P, Yamashita Y, RIETE investigators, et al. Derivation and validation of a clinical prediction rule for thrombolysis-associated major bleeding in patients with acute pulmonary embolism: the BACS score. Eur Respir J. 2020. https://doi.org/10.1183/13993003.02336-2020.

25. Badescu MC, Ciocoiu M, Badulescu OV, Vladeanu MC, Bojan IB, Vlad CE, et al. Prediction of bleeding events using the VTE-BLEED risk score in patients with venous thromboembolism receiving anticoagulant therapy (review). Exp Ther Med. 2021;22(5):1344. https://doi.org/10.3892/etm.2021.10779. (**Epub 2021 Sep 22**).
26. Klok FA, Hösel V, Clemens A, Yollo WD, Tilke C, Schulman S, et al. Prediction of bleeding events in patients with venous thromboembolism on stable anticoagulation treatment. Eur Respir J. 2016;48(5):1369–76. https://doi.org/10.1183/13993003.00280-2016. (**Epub 2016 Jul 28**).
27. Klok FA, Barco S, Konstantinides SV. Evaluation of VTE-BLEED for predicting intracranial or fatal bleeding in stable anticoagulated patients with venous thromboembolism. Eur Respir J. 2018;51(4):1800077. https://doi.org/10.1183/13993003.00077-2018.

Authors and Affiliations

María Carmen Viñuela-Benéitez[1] · Claudia Iglesias Pérez[2] · Laura Ortega Morán[3] · Ignacio García Escobar[4] · Diego Cacho Lavín[5] · Rut Porta i Balanyà[6] · Silvia García Adrián[7] · Marta Carmona Campos[8] · Gretel Benítez López[9] · José Antonio Santiago Crespo[10] · Miriam Lobo de Mena[11] · Javier Pérez Altozano[12] · Enrique Gallardo Díaz[13] · Julia Tejerina Peces[14] · Pilar Ochoa Rivas[15] · María José Ortiz Morales[16] · Victoria Eugenia Castellón Rubio[17] · Carmen Díez Pedroche[18] · María Rosales Sueiro[19] · Felipe Gonçalves[20] · Manuel Sánchez-Cánovas[21] · Miguel Ángel Ruiz[22] · José Muñoz-Langa[23] · Pedro Pérez Segura[14] · Eva Martínez de Castro[5] · Alberto Carmona-Bayonas[21] · Paula Jiménez-Fonseca[2] · Andrés Jesús Muñoz Martín[3]

✉ Andrés Jesús Muñoz Martín
andmun13@ucm.es

[1] Department of Medicine, Faculty of Medicine, Universidad Complutense, Madrid, Spain

[2] Medical Oncology Department, Hospital Universitario Central de Asturias, ISPA, Oviedo, Spain

[3] Department of Medical Oncology, Instituto de Investigación Sanitaria Gregorio Marañón, Universidad Complutense, Madrid, Spain

[4] Department of Medical Oncology, Hospital Universitario de Toledo, Toledo, Spain

[5] Department of Medical Oncology, Hospital Universitario Marqués de Valdecilla (Idival), Santander, Cantabria, Spain

[6] Department of Medical Oncology, Institut Català d'Oncologia (ICO), Hospital Universitari de Girona Doctor Josep Trueta, Girona, Spain

[7] Department of Medical Oncology, Hospital Universitario de Móstoles, Madrid, Spain

[8] Department of Medical Oncology, Hospital Clínico Universitario de Santiago de Compostela, A Coruña, Spain

[9] Department of Medical Oncology, C.H.U. Insular-Materno Infantil de Gran Canaria, Las Palmas, Spain

[10] Department of Medical Oncology, Hospital Virgen de La Luz, Cuenca, Spain

[11] Department of Medical Oncology, Consorcio Hospital General Universitario de Valencia, Valencia, Spain

[12] Department of Medical Oncology, IMED Hospital Elche, Alicante, Spain

[13] Department of Oncology, Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Parc Taulí Hospital Universitari, Universitat Autònoma de Barcelona, Sabadell, Spain

[14] Department of Medical Oncology, Hospital Clínico San Carlos (Madrid), Instituto de Investigación Clínico San Carlos (IdISSC), Madrid, Spain

[15] Department of Medical Oncology, Hospital Central de la Defensa Gómez Ulla, Madrid, Spain

[16] Department of Medical Oncology, Hospital Universitario Reina Sofía, Córdoba, Spain

[17] Department of Medical Oncology, Hospital Universitario Torrecárdenas, Almería, Spain

[18] Department of Internal Medicine, Hospital 12 de Octubre, Universidad Complutense, Madrid, Spain

[19] Department of Immunohemotherapy, Portuguese Institute of Oncology of Porto (IPO Porto), Porto, Portugal

[20] Department of Immunohemotherapy, Portuguese Institute of Oncology of Lisbon Francisco Gentil (IPO Lisboa), Lisboa, Portugal

[21] Department of Medical Oncology, Hospital Universitario José Maria Morales Meseguer, Murcia, Spain

[22] School of Psychology, Universidad Autónoma de Madrid, Madrid, Spain

[23] Department of Medical Oncology, Hospital Arnau de Villanova de Valencia, Valencia, Spain