Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Clinical Phenotypes and Prognosis in Patients with Venous Thromboembolism and Cancer: A Cluster Analysis

Arturo D. Mora
Fundación Valle del Lili
Juan Pablo Gómez-Arellano
Universidad ICESI
David Fernando Narvaez-De Los Rios
Fundación Valle del Lili
Manuela Estrada-Schweineberg
Universidad ICESI
Andrés Domínguez-Durán
Universidad ICESI
Valeria Azcarate-Rodríguez
Fundación Valle del Lili
Juan Andres Muñoz
Fundación Valle del Lili
Diana C. Carrillo-Gómez
Fundación Valle del Lili
Julián David Yara
Fundación Valle del Lili
Hoover León-Giraldo
Fundación Valle del Lili
Frank Douglas Cañón-Estrada
Fundación Valle del Lili
Mario Miguel Barbosa-Rengifo
Fundación Valle del Lili
Jessica Largo-Ocampo
Fundación Valle del Lili
Juan Esteban Gómez-Mesa
juan.gomez.me@fvl.org.co

Fundación Valle del Lili

Research Article

Keywords:

Posted Date: July 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10452942/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Thrombosis Journal on November 15th, 2026. See the published version at https://doi.org/10.1186/s12959-026-00940-y.

Abstract

Background

Risk stratification in patients with venous thromboembolism (VTE) and cancer is predominantly tumor-centered, although outcomes are also influenced by patient clinical characteristics. Whether non-oncologic clinical profiles identify prognostic subgroups independent of tumor-related thrombotic risk remains insufficiently explored.

Methods

We conducted a retrospective cohort study of patients with venous thromboembolism (VTE) and current or prior malignancy treated at a tertiary referral center in Colombia between January 2011 and January 2025. Clinical phenotypes were identified using multiple correspondence analysis followed by hierarchical clustering (Ward's method), excluding tumor site and oncologic treatments from the clustering process. Overall survival (OS) was estimated using Kaplan–Meier analysis. Cox proportional hazards models were used to evaluate the association between clinical phenotypes and mortality, adjusting for tumor-related thrombotic risk categories.

Results

A total of 654 patients were included (median age 66.6 years; 68.8% female; 72.9% with active cancer at VTE diagnosis). Median OS for the entire cohort was 3.7 years (95% CI 2.8–5.4). Three clinical phenotypes were identified: (1) younger patients with active cancer (n = 367), (2) older patients with recent surgical exposure (n = 211), and (3) older patients with multimorbidity (n = 76). Median OS differed across phenotypes (2.4, 6.2, and 3.8 years, respectively). Low-molecular-weight heparin was the predominant anticoagulant (76.5%). In multivariable analysis, both clinical phenotype and tumor-related thrombotic risk were independently associated with OS.

Conclusions

Among patients with VTE and malignancy, clinical phenotypes defined by non-oncologic characteristics identify prognostic groups beyond tumor-related thrombotic risk. These findings highlight the relevance of patient clinical profile in outcome heterogeneity within this population.

INTRODUCTION

Venous thromboembolism (VTE) is a frequent and clinically significant complication in patients with cancer. It is estimated that 20%–30% of all newly diagnosed VTE events occur in individuals with malignancy, facing a nearly ninefold higher risk compared with the general population (1–3). Importantly,

thromboembolic events, commonly referred to as cancer-associated thrombosis, represent the second leading cause of death in patients with cancer, highlighting their substantial impact on survival (4, 5).

Cancer-associated VTE arises from a complex interplay between tumor-related factors, oncologic treatments, and patient-specific characteristics. Advanced or metastatic disease and cytotoxic chemotherapy have been consistently associated with increased thrombotic risk (6–8). Certain malignancies, such as pancreatic and gastrointestinal cancers, exhibit particularly high thrombogenic potential. Additional factors-including multimorbidity, hospitalization, and recent oncologic surgery-further amplify thrombotic vulnerability in this population (9, 10).

Despite the high burden of VTE in this population, current guidelines do not recommend routine pharmacological thromboprophylaxis for all ambulatory cancer patients due to concerns regarding bleeding risk and the heterogeneity of individual risk profiles (11). Risk prediction models, such as the Khorana score, have facilitated targeted prophylaxis in selected populations; however, their real-world applicability and predictive performance remain variable (12, 13).

Current prognostic assessment in this population largely relies on tumor-related risk stratification (12, 13). However, clinical outcomes in patients with VTE and malignancy may also be influenced by patient-level characteristics independent of tumor biology, such as comorbidity burden, recent clinical exposures, and overall clinical status (9, 10). As a result, patients with similar tumor types often experience markedly different clinical courses, suggesting the presence of underlying clinical heterogeneity not captured by tumor-based models alone (14).

Cluster-based analytical approaches have demonstrated value in identifying clinically meaningful phenotypes with distinct prognostic trajectories in several cardiovascular and thrombotic conditions, including atrial fibrillation, heart failure, and VTE (14–17). Applying a similar framework to VTE in patients with cancer may help characterize patient profiles beyond oncologic features and improve understanding of outcome variability in this population.

Therefore, the present study aimed to identify clinical phenotypes among patients with VTE and malignancy based on non-oncologic characteristics, and to evaluate whether these phenotypes are associated with survival outcomes beyond tumor-related thrombotic risk.

METHODS

Study design and population

This retrospective observational study was conducted in a tertiary referral center located in Cali, Colombia. The study population comprised patients evaluated at the outpatient Anticoagulation Clinic of the Cardiology Department between January 2011 and January 2025.

Patients were eligible for inclusion if they had an objectively confirmed diagnosis of VTE, including deep vein thrombosis (DVT), pulmonary embolism (PE), or venous thrombosis at unusual sites, such as abdominal or cerebral veins.

Eligible patients were required to have either active cancer or a prior history of cancer at the time of VTE diagnosis. Patients whose only malignancy was basal cell or squamous cell carcinoma of the skin were excluded. Additionally, patients with incomplete or insufficient clinical records were excluded from the analysis (Fig. 1).

Active cancer was defined as a diagnosis of malignancy with ongoing antineoplastic or palliative treatment documented at the time of clinical evaluation. Patients with documented remission, inactive disease, or cure were classified as having a history of cancer.

Clinical variables and outcomes

Clinical and demographic data were extracted from electronic medical records using a standardized digital questionnaire. Collected variables included oncologic characteristics, non-oncologic comorbidities, VTE-related characteristics, and anticoagulant treatment. Multimorbidity was defined as the presence of two or more comorbid conditions.

OS was the primary outcome. Bleeding and recurrent VTE were considered secondary exploratory outcomes. Vital status was ascertained through medical records and official records from the Colombian Administrator of Resources of the General Social Security System in Health (Administradora de los Recursos del Sistema General de Seguridad Social en Salud, ADRES).

OS was defined as the time from the index VTE event to death from any cause or last follow-up. The index VTE event was selected as a common clinical milestone across patients, representing the point at which thrombosis was clinically recognized and anticoagulation management initiated. Because VTE frequently occurs at different stages of the cancer trajectory, survival analyses were interpreted as post-VTE outcomes rather than cancer-specific prognosis. Accordingly, the objective was not to evaluate tumor-related mortality, but to assess outcome heterogeneity among patients with VTE and malignancy after a shared clinical event.

Bleeding complications were classified according to the criteria of the International Society on Thrombosis and Haemostasis (ISTH) (18). Major bleeding was defined as any of the following: (1) a decrease in hemoglobin level of ≥2 g/dL; (2) transfusion of ≥2 units of packed red blood cells; or (3) bleeding occurring at a critical site, including intracranial, intraspinal, intraocular, pericardial, intra-articular, intramuscular with compartment syndrome, or retroperitoneal bleeding. Clinically relevant non-major bleeding (CRNMB) was defined as overt bleeding not meeting the criteria for major bleeding but fulfilling at least one of the following conditions: [1] requiring non-surgical medical intervention; [2] leading to hospitalization or an increased level of care; or [3] prompting an unscheduled in-person

medical evaluation (19). Minor bleeding was defined as overt bleeding events that did not require significant clinical intervention or changes in management.

Extended duration of anticoagulation was defined as anticoagulant treatment lasting more than six months.

Statistical analysis

An initial descriptive analysis was performed. Continuous variables were summarized as mean (standard deviation [SD]) or median (interquartile range [IQR]) according to distribution assessed using the Shapiro–Wilk test. Categorical variables were reported as absolute frequencies and percentages. Differences in baseline characteristics were evaluated using the Kruskal–Wallis test for continuous variables and the chi-square or Fisher's exact test, as appropriate, for categorical variables.

Identification of clinical phenotypes:

Sixteen variables reflecting patient clinical characteristics, comorbidities, and recent clinical exposures related to VTE were selected for the clustering procedure (Table S1). Variables directly related to tumor biology, including primary tumor site and oncologic treatments, were intentionally excluded to ensure that the derived groups represented patient clinical phenotypes rather than oncologic categories. Inclusion of tumor-related variables in unsupervised clustering would likely result in partitions driven primarily by tumor type, a known major determinant of prognosis, thereby limiting identification of non-oncologic profiles.

Multiple correspondence analysis (MCA) was performed, and hierarchical agglomerative clustering using Ward's minimum-variance method was applied to the MCA-derived factor scores to identify phenotypes. The optimal number of clusters was determined by jointly evaluating average silhouette width across solutions ranging from k = 2 to k = 7 and clinical interpretability (Figure S1). Although the two-cluster solution yielded the highest silhouette width, the three-cluster solution provided greater clinical resolution and was therefore selected, consistent with recommendations emphasizing domain-specific interpretability alongside quantitative metrics in clinical phenotyping studies (14–16).

Survival analysis:

Survival curves were estimated using the Kaplan–Meier method and compared using the log-rank test. Survival analyses were administratively truncated at 10 years due to the limited number of individuals remaining at risk beyond this time point, which could produce unstable estimates and wide confidence intervals.

Because Cluster 3 was substantially smaller and showed a higher proportion of censored observations, an additional sensitivity analysis was performed restricting the cohort to Clusters 1 and 2. This analysis

aimed to evaluate whether the association between clinical phenotype and mortality persisted independently of the smallest cluster and to assess the stability of the primary model estimates.

Multivariable analysis:

Cox proportional hazards regression was used to estimate hazard ratios (HRs) for mortality. Cluster membership was considered the primary exposure variable. The model was adjusted for tumor-related thrombotic risk, categorized into three groups according to Khorana et al. (12): standard risk (breast, prostate, lower gastrointestinal, osteomuscular, head and neck, neuroendocrine, and other tumors), high risk (gynecological, lung, hepatobiliary, hematological, urologic, or cerebral tumors), and very high risk (pancreatic and upper gastrointestinal tumors). Tumor-related thrombotic risk was included to evaluate whether the association between clinical phenotype and mortality was independent of tumor-related risk. Variables used to derive the clusters were not simultaneously included in the model to avoid overadjustment. The proportional hazards assumption was assessed using Schoenfeld residuals.

All statistical analyses were conducted using R (version 2023.12.0+369), adopting a significance threshold of $p < 0.05$. Figures were generated using R, Microsoft Excel (v2311), Lucidchart, and BioRender.

Ethical consideration

The study was approved by the Biomedical Ethics and Research Committee of Fundación Valle del Lili (Protocol No. 2018.1297) in accordance with Resolution 8430 of 1993 and the Declaration of Helsinki. Because the study used retrospective clinical data and did not modify patient management, it was classified as minimal risk and granted a waiver of informed consent.

The study also complied with Colombian Habeas Data regulations (Statutory Law 1581 of 2012 and Regulatory Decree 1377 of 2013) regarding confidentiality, data protection, and appropriate handling of sensitive health information.

RESULTS

General population characteristics

A total of 654 patients with VTE and malignancy were included (Figure 1). Median age was 66.6 years (IQR 56.6–76.9), and 68.8% were female. Active cancer at the time of the index VTE was present in 72.9%. The most frequent primary malignancies were breast (23.1%), gastrointestinal (24.7%), and gynecologic tumors (15.7%). Hypertension was the most prevalent comorbidity (41.4%), and 34.3% of patients met criteria for multimorbidity (≥2 comorbidities) (Table 1, 2).

Table 1. Clinical and oncological characteristics

	Overall, N = 654	Cluster1, N = 367	Cluster 2, N = 211	Cluster 3 N = 76	P value
Gender					<0.001
Female	450 (68.8%)	286 (77.9%)	118 (55.9%)	46 (60.5%)	
Male	204 (31.2%)	81 (22.1%)	93 (44.1%)	30 (39.5%)	
Age*	66.6 (56.6-76.9)	59.3 (48.9-65.6)	75.9 (69.3-81.3)	78.5 (71.5-85.5)	<0.001
Age group					<0.001
Over 65 y/o	355 (54.3%)	99 (27.0%)	189 (89.6%)	67 (88.2%)	
50 to 65 y/o	197 (30.1%)	169 (46.0%)	21 (10.0%)	7 (9.2%)	
Under 50 y/o	102 (15.6%)	99 (27.0%)	1 (0.5%)	2 (2.6%)	
Active cancer	477 (72.9%)	300 (81.7%)	130 (61.6%)	47 (61.8%)	<0.001
Primary tumor site					
Brain or spinal cord	11 (1.7%)	8 (2.2%)	2 (0.9%)	1 (1.3%)	
Head and neck	28 (4.3%)	8 (2.2%)	14 (6.6%)	6 (7.9%)	
Upper GI tract	42 (6.4%)	22 (6.0%)	17 (8.1%)	3 (3.9%)	
Liver / biliary tract	16 (2.4%)	6 (1.6%)	7 (3.3%)	3 (3.9%)	
Pancreas	30 (4.6%)	16 (4.4%)	11 (5.2%)	3 (3.9%)	
Lower GI tract	74 (11.3%)	37 (10.1%)	24 (11.4%)	13 (17.1%)	
Breast	151 (23.1%)	92 (25.1%)	38 (18.0%)	21 (27.6%)	
Gynecologic	103 (15.7%)	79 (21.5%)	19 (9.0%)	5 (6.6%)	
Lung	32 (4.9%)	18 (4.9%)	10 (4.7%)	4 (5.3%)	
Prostate	52 (8.0%)	11 (3.0%)	32 (15.2%)	9 (11.8%)	
Urologic	46 (7.0%)	27 (7.4%)	17 (8.1%)	2 (2.6%)	
Hematologic	46 (7.0%)	28 (7.6%)	12 (5.7%)	6 (7.9%)	
Soft tissue and bone	14 (2.1%)	8 (2.2%)	6 (2.8%)	0 (0.0%)	
Neuroendocrine or other	9 (1.4%)	7 (1.9%)	2 (0.9%)	0 (0.0%)	
Cancer treatment					
Chemotherapy	461 (70.9%)	292 (80.2%)	126 (60.0%)	43 (56.6%)	<0.001

Surgery	365 (56.2%)	185 (50.8%)	132 (62.9%)	48 (63.2%)	0.008
Radiotherapy	213 (32.6%)	120 (32.8%)	68 (32.2%)	25 (32.9%)	>0.9
Palliative care	164 (25.3%)	95 (26.2%)	50 (23.9%)	19 (25.0%)	0.8

*Data are presented as n (%), unless otherwise specified. *Age is reported as median (interquartile range). Gastrointestinal cancers were classified as upper and lower gastrointestinal tract malignancies. Pancreatic and hepatobiliary cancers were analyzed as separate categories. Abbreviations: N, number of patients; y/o, years old; GI, gastrointestinal; p, P value.*

Identification of clinical phenotypes

Multiple correspondence analysis followed by hierarchical clustering identified three distinct clinical phenotypes (Figure 2, Table 1 and 2). Although the two-cluster solution showed the highest silhouette width, the three-cluster solution provided greater clinical resolution and interpretability and was therefore selected:

- **Cluster 1 - Younger patients with active cancer** (n=367) comprised predominantly younger patients (median 59.3 years) with a high prevalence of active cancer (81.7%) and low comorbidity burden.
- **Cluster 2 - Older patients with recent surgical exposure** (n=211) included older individuals (median 75.9 years) exposed to recent surgery and with intermediate comorbidity burden.
- **Cluster 3 - Older patients with multimorbidity** (n=76) consisted of older patients (median 78.5 years) with marked cardiovascular multimorbidity and the highest HAS-BLED risk profile.

Table 2. Non-oncological comorbidities

	Overall, N = 654	Cluster 1, N = 367	Cluster 2, N = 211	Cluster 3, N = 76	p value
Cardiovascular comorbidities					
Diabetes mellitus	92 (14.1%)	9 (2.5%)	68 (32.2%)	15 (19.7%)	<0.001
Obesity	63 (14.6%)	31 (12.1%)	22 (16.2%)	10 (25.6%)	0.067
Atherosclerotic disease	47 (7.2%)	6 (1.6%)	19 (9.0%)	22 (28.9%)	<0.001
Cerebrovascular disease	37 (5.7%)	8 (2.2%)	1 (0.5%)	28 (36.8%)	<0.001
Heart failure	35 (5.4%)	5 (1.4%)	3 (1.4%)	27 (35.5%)	<0.001
Atrial fibrillation	35 (5.4%)	3 (0.8%)	2 (0.9%)	30 (39.5%)	<0.001
Valvular disease	30 (4.6%)	9 (2.5%)	4 (1.9%)	17 (22.4%)	<0.001
Arterial hypertension	271 (41.4%)	52 (14.2%)	159 (75.4%)	60 (78.9%)	<0.001
Other comorbidities					
Chronic kidney disease	43 (6.6%)	14 (3.8%)	26 (12.3%)	3 (3.9%)	<0.001
Chronic Obstructive Pulmonary Disease	24 (3.7%)	3 (0.8%)	0 (0.0%)	21 (27.6%)	<0.001
History of previous VTE	47 (7.2%)	21 (5.7%)	15 (7.1%)	11 (14.5%)	0.027
HASBLED ≥ 3	110 (16.8%)	2 (0.5%)	70 (33.2%)	38 (50.0%)	<0.001
Presence of 2 or more comorbidities	224 (34.3%)	38 (10.4%)	114 (54.0%)	72 (94.7%)	<0.001

Abbreviations: N, number of patients; p, P value; VTE: venous thromboembolism .

VTE presentation and anticoagulation management

DVT alone occurred in 48.3%, PE alone in 33.5%, concomitant PE/DVT in 9.6%, and unusual-site DVT in 12.4% of patients. Recent surgery (last 3 months) was most frequent in Cluster 2 (28.0%), whereas recent central venous catheter use (last 3 months) was more common in Cluster 1 (30.8%). Extended anticoagulation (>6 months) was recommended in 84.2% of cases without differences across clusters. Low-molecular-weight heparin (LMWH) was the predominant anticoagulant (76.5%) (Table 3).

Table 3. VTE characteristics and anticoagulant treatment

	Overall, N = 654	Cluster 1, N = 367	Cluster 2, N = 211	Cluster 3, N = 76	P value
VTE					
• DVT alone	316 (48.3%)	201 (54.8%)	90 (42.7%)	25 (32.9%)	<0.001
• PE alone	219 (33.5%)	99 (27.0%)	84 (39.8%)	36 (47.4%)	
• PE/DVT	63 (9.6%)	35 (9.5%)	15 (7.1%)	13 (17.1%)	
• Unusual site VTE	81 (12.4%)	45 (12.3%)	31 (14.7%)	5 (6.6%)	0.2
Transient risk factors					
• Hospitalization	236 (36.1%)	122 (33.2%)	84 (39.8%)	30 (39.5%)	0.2
• Surgery	115 (17.6%)	48 (13.1%)	59 (28.0%)	8 (10.5%)	<0.001
• Central venous catheter	141 (21.6%)	113 (30.8%)	22 (10.4%)	6 (7.9%)	<0.001
Recommended duration of anticoagulation					
• 3 months	60 (9.2%)	38 (10.4%)	19 (9.1%)	3 (3.9%)	0.5
• 6 months	43 (6.6%)	24 (6.5%)	13 (6.2%)	6 (7.9%)	
• Extended	549 (84.2%)	305 (83.1%)	177 (84.7%)	67 (88.2%)	
Anticoagulant					
• Apixaban	36 (5.5%)	16 (4.4%)	16 (7.6%)	4 (5.3%)	<0.001
• Dabigatran	6 (0.9%)	0 (0.0%)	5 (2.4%)	1 (1.3%)	
• Low-molecular weight heparin	500 (76.5%)	313 (85.3%)	148 (70.1%)	39 (51.3%)	
• Rivaroxaban	54 (8.3%)	20 (5.4%)	20 (9.5%)	14 (18.4%)	
• Warfarin	58 (8.9%)	18 (4.9%)	22 (10.4%)	18 (23.7%)	

VTE: venous thromboembolism, DVT: Deep vein thrombosis, PE: Pulmonary embolism. Transient risk factors refers to events that ocurred in the 3 months prior to the VTE event

Survival analysis

Overall survival

Median OS for the entire cohort was 3.7 years (95% CI 2.8–5.4). The Kaplan–Meier curve showed a higher event rate during early follow-up followed by stabilization of survival probabilities over time (Figure 3). Survival analyses were administratively truncated at 10 years due to the limited number of individuals remaining at risk beyond this time frame, which could yield unstable estimates. Additional analyses addressing censoring are provided in the Supplementary Material (Table S2 and Figure S2).

Survival stratified by clinical phenotypes

Survival differed across clinical phenotypes (log-rank p=0.044) (Figure 4). Median OS was 2.4 years in Cluster 1, 6.2 years in Cluster 2, and 3.8 years in Cluster 3. In Cluster 3, the upper confidence bound for the median survival could not be estimated due to the proportion of censored observations.

Secondary outcomes

Major bleeding occurred in 3.8% and recurrent VTE in 3.8% of patients, without significant differences across clusters (Table 4).

Multivariable Cox regression

In the Cox proportional hazards model adjusted for tumor-related thrombotic risk, Cluster 2 was used as the reference category. Compared with Cluster 2, Cluster 1 was associated with higher mortality (HR 1.30; 95% CI 1.02–1.65; p=0.032). Cluster 3 showed no statistically significant difference in mortality risk relative to Cluster 2 (HR 1.07; 95% CI 0.74–1.54; p=0.731) (Figure 5).

The proportional hazards assumption was not violated (global Schoenfeld test p=0.070). Model diagnostics are provided in the Supplementary Material

Adjusted survival analysis

Survival curves derived from the Cox model adjusted for tumor-related thrombotic risk demonstrated persistent separation between phenotypes (Figure 6). Differences were primarily observed during the early follow-up period, with increasing overlap at later time due to censoring. The overall pattern of adjusted survival was consistent with the hazard ratios observed in the multivariable analysis.

Table 4. Complications associated with anticoagulant treatment

Variable	Overall (N = 654)	Cluster 1 (N = 367)	Cluster 2 (N = 211)	Cluster 3 (N = 76)	p-value
Bleeding	43 (6.6%)	19 (5.2%)	14 (6.6%)	10 (13.2%)	0.049
Classification					0.052
Major	20 (46.5%)	6 (31.6%)	6 (42.9%)	8 (80.0%)	
CRNMB	4 (9.3%)	1 (5.3%)	2 (14.3%)	1 (10.0%)	
Minor	19 (44.2%)	12 (63.2%)	6 (42.9%)	1 (10.0%)	
Location of bleeding					
Intracranial	2 (4.9%)	0 (0.0%)	0 (0.0%)	2 (20.0%)	0.055
Gastrointestinal	14 (32.6%)	3 (15.8%)	6 (42.9%)	5 (50.0%)	0.094
Genitourinary	13 (31.0%)	5 (27.8%)	6 (42.9%)	2 (20.0%)	0.50
Gynecological	8 (19.0%)	6 (33.3%)	2 (14.3%)	0 (0.0%)	0.11
Oral	3 (7.1%)	2 (11.1%)	0 (0.0%)	1 (10.0%)	0.50
Soft tissues	3 (7.1%)	2 (11.1%)	1 (7.1%)	0 (0.0%)	0.80
Recurrent VTE	25 (3.8%)	12 (3.3%)	9 (4.3%)	4 (5.3%)	0.65

Percentages were calculated among patients with bleeding events. Abbreviations: CRNMB: Clinically relevant non-major bleeding Data are presented as n (%). Percentages were calculated using available cases. p-values were obtained using Fisher's exact test or Pearson's chi-squared test, as appropriate.

DISCUSSION

In this cohort of 654 patients with VTE and malignancy, we identified three clinically distinct phenotypes using an unsupervised clustering approach based on non-oncologic characteristics. These phenotypes were associated with differences in OS following the index VTE event. Importantly, the association between clinical phenotype and mortality persisted after adjustment for tumor-related thrombotic risk, suggesting that patient-level clinical profile contributes to prognostic heterogeneity beyond tumor-related factors alone.

Mortality and prognostic relevance of VTE in patients with cancer

VTE in patients with cancer has long been recognized as a marker of adverse prognosis rather than an isolated vascular complication. Early population-based studies demonstrated substantially reduced survival among patients diagnosed with cancer in the context of VTE. This condition reflects advanced

disease burden and host vulnerability rather than a purely thrombotic event (20). Subsequent Scandinavian cohorts reported persistently elevated mortality following cancer-associated VTE across cancer types, despite improvements in oncologic care, highlighting a particularly high-risk period in the months after the thrombotic event (21,22). Real-world registry data from RIETE and GARFIELD-VTE have consistently identified active cancer as one of the strongest predictors of early mortality after VTE (23,24).

In our cohort, median OS after the index VTE event was 3.7 years, with a marked concentration of events during early follow-up. These findings are consistent with prior literature and reinforce the concept that VTE in the setting of malignancy identifies a clinically vulnerable population in whom prognosis is driven by both tumor-related and patient-level factors.

Potential clinical implication

This cluster-based analysis illustrates the presence of clinically distinct phenotypes characterized by differences in comorbidity burden, cancer activity, and bleeding risk profile. Notably, Cluster 1 comprised younger individuals but demonstrated shorter OS, whereas Cluster 2 exhibited longer survival despite older age, suggesting that prognostic heterogeneity extends beyond chronological age alone.

In addition to clinical phenotype, tumor-related thrombotic risk remains a key determinant of prognosis, as reflected in the Khorana model and subsequent validations (12,19,25). Consistent with prior literature, patients with very high-risk malignancies such as pancreatic and upper gastrointestinal cancers experienced poorer survival compared with those with lower-risk tumors such as breast or prostate cancer (26,27). These findings support the concept that prognosis in VTE in patients with cancer reflects the combined influence of tumor biology and patient-level characteristics.

Although our study did not evaluate cause-specific mortality or treatment effects, identification of phenotypic profiles associated with differential survival may have implications for clinical assessment. Recognition of higher-risk profiles could support closer follow-up and more individualized evaluation of anticoagulation management, particularly in patients with concomitant bleeding risk or complex comorbidity. Additionally, the distribution of recent surgical exposure across phenotypes highlights the importance of contextualizing VTE events within perioperative oncology care settings (28,29).

Taken together, these findings suggest that integrating clinical phenotype with tumor-related risk may contribute to a more nuanced understanding of prognosis in this population.

Anticoagulant Treatment in Patients with VTE and Cancer

In our cohort, LMWH was the most frequently prescribed anticoagulant, reflecting contemporary guideline-based practice during much of the study period. Current international recommendations endorse the use of LMWH and selected direct oral anticoagulants for the treatment of VTE in patients

with cancer, with therapeutic choice guided by tumor type, bleeding risk, drug–drug interactions, and patient-specific factors (11,13,28).

Although our study did not evaluate comparative effectiveness or safety across different anticoagulant regimens, the distribution of anticoagulant use across phenotypes provides contextual insight into real-world management. The integration of thrombotic risk, tumor biology, and bleeding profile-elements partially captured by our phenotypic classification-remains central to individualized anticoagulation decisions (13,28). Thus, clinical phenotyping may complement tumor-related risk stratification when considering anticoagulant strategies in patients with malignancy.

Limitations and strengths

This study has limitations inherent to its retrospective observational design, which precludes causal inference. The single-center setting may limit external generalizability, and external validation in independent cohorts will be necessary to confirm reproducibility of the identified phenotypes. In addition, detailed oncologic variables such as tumor stage, molecular subtype, and timing of remission were not systematically available and therefore could not be incorporated into the clustering procedure. As a result, residual confounding by unmeasured tumor severity cannot be excluded.

Two additional limitations warrant consideration. First, Cluster 3 comprised a smaller number of patients, which may have limited statistical precision for estimates within this subgroup. However, sensitivity analyses restricted to Clusters 1 and 2 yielded consistent findings, supporting the stability of the primary results. Second, cause-specific mortality was not available, precluding differentiation between cancer-related and thrombotic-related deaths.

Despite these limitations, the study has notable strengths. The cohort included a substantial number of patients with long-term follow-up extending up to 10 years, allowing evaluation of clinically relevant post-VTE outcomes. Although censoring increased over time, analyses were driven primarily by events occurring during early follow-up, when risk was highest. The application of an unsupervised clustering approach enabled data-driven identification of clinically interpretable phenotypes without imposing predefined categories. Finally, the inclusion of a broad spectrum of malignancies and comorbid conditions within a real-world tertiary-care population enhances the clinical relevance of the findings.

Further validation in independent, multicenter cohorts will be essential to confirm the reproducibility and prognostic consistency of the identified phenotypes. Prospective studies incorporating detailed oncologic staging and longitudinal assessment of cancer activity may clarify the interaction between tumor biology and patient-level clinical phenotype and determine whether phenotype-informed strategies meaningfully refine risk assessment in patients with VTE and cancer.

CONCLUSION

VTE in patients with cancer represents a clinically heterogeneous condition characterized by substantial variability in patient profiles and outcomes. Using an unsupervised clustering approach based on non-oncologic characteristics, we identified three clinically distinct phenotypes associated with differences in OS. The association between clinical phenotype and mortality persisted after adjustment for tumor-related thrombotic risk, suggesting that patient-level factors contribute to prognostic heterogeneity beyond tumor biology alone.

These findings support the relevance of clinical phenotyping as a complementary framework to tumor-based risk stratification in VTE in patients with cancer. Prospective validation is warranted to determine whether phenotype-informed approaches can enhance prognostic assessment and clinical management.

Declarations

Acknowledgements

We would like to thank We also extend our gratitude to the Research Center at Fundación Valle del Lili for its institutional support.

Use of artificial intelligence:

Artificial intelligence programs (large language models) were used for proofreading and spell check.

Conflicts of Interest:

The authors declare no competing interests.

Data Availability Statement:

Due to the data containing personal, sensitive data, the data used in this study can be made available upon reasonable request to researchers who meet the criteria for access to confidential data. Requests should be submitted in writing to the Research Ethics Committee of Fundación Valle del Lili: Comité de Ética en Investigación Biomédica, Fundación Valle del Lili Email: cei@fvl.org.co

Competing interests:

The authors have declared that no competing interests exist.

Funding

This research was financially supported by the Centro de Investigaciones Clínicas at Fundación Valle de Lili. The findings and conclusions presented in this article are solely those of the authors.

Authorship

Conceptualization: J.E.G.M, A.D.M, J.P.G.A, D.F.N.D, M.E.S, A.D.D, D.C.C.G, J.D.Y, H.L.G, M.M.B.R; Data curation, J.E.G.M, A.D.M, H.L.G, J.P.G.A, M.E.S, M.M.B.R.; Formal analysis H.L.G, M.M.B.R, J.L.O; Funding acquisition, J.E.G.M.; Investigation A.D.M, J.E.G.M, J.P.G.A, D.F.N.D, M.E.S, A.D.D, D.C.C.G, J.D.Y, H.L.G and M.M.B.R.; Methodology H.L.G, M.M.B.R, A.D.M. and J.E.G.M.; Project administration: J.E.G.M.; Resources: J.E.G.M; Software: H.L.G, M.M.B.R, J.L.O; Supervision: J.E.G.M, A.D.M, D.F.N.D, H.L.G, and M.M.B.R.; Validation: J.E.G.M, A.D.M, H.L.G and M.M.B.R; Visualization: A.D.M, M.M.B.R, J.L.O and V.A.R; Writing-original draft: J.E.G.M, A.D.M, J.P.G.A, M.E.S, D.F.N.D, A.D.D, D.C.C.G, J.D.Y, H.L.G, F.D.M, M.M.B.R, J.A.M; Writing-review and editing:J.E.G.M, A.D.M, J.P.G.A, M.E.S, D.F.N.D, F.D.M, A.D.D, D.C.C.G, J.D.Y, H.L.G, M.M.B.R, J.A.M. All authors have read and agreed to the published version of the manuscript.

References

1. Khorana AA, Connolly GC. Assessing Risk of Venous Thromboembolism in the Patient With Cancer. J Clin Oncol. 2009;27(29):4839–47. 10.1200/JCO.2009.22.3271.
2. Timp JF, Braekkan SK, Versteeg HH, Cannegieter SC. Epidemiology of cancer-associated venous thrombosis. Blood. 2013;122(10):1712–23. 10.1182/blood-2013-04-460121.
3. Walker AJ, Card TR, West J, Crooks C, Grainge MJ. Incidence of venous thromboembolism in patients with cancer – A cohort study using linked United Kingdom databases. Eur J Cancer. 2013;49(6):1404–13. 10.1016/j.ejca.2012.10.021.
4. Mulder FI, Horváth-Puhó E, Van Es N, Van Laarhoven HWM, Pedersen L, Moik F, et al. Venous thromboembolism in cancer patients: a population-based cohort study. Blood. 2021;137(14):1959–69. 10.1182/blood.2020007338.
5. 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. 10.1111/j.1538-7836.2007.02374.x.
6. O'Connell C, Escalante CP, Goldhaber SZ, McBane R, Connors JM, Raskob GE. Treatment of Cancer-Associated Venous Thromboembolism with Low-Molecular-Weight Heparin or Direct Oral Anticoagulants: Patient Selection, Controversies, and Caveats. Oncologist. 2021;26(1):e8–16. 10.1002/onco.13584.
7. Moik F, Ay C, Pabinger I. Risk prediction for cancer-associated thrombosis in ambulatory patients with cancer: past, present and future. Thromb Res. 2020;191:S3–11. 10.1016/S0049-3848(20)30389-3.
8. Abdol Razak N, Jones G, Bhandari M, Berndt M, Metharom P. Cancer-Associated Thrombosis: An Overview of Mechanisms, Risk Factors, and Treatment. Cancers. 2018;10(10):380. 10.3390/cancers10100380.
9. Tsantes AG, Petrou E, Tsante KA, Sokou R, Frantzeskaki F, Domouchtsidou A, et al. Cancer-Associated Thrombosis: Pathophysiology, Laboratory Assessment, and Current Guidelines. Cancers. 2024;16(11):2082. 10.3390/cancers16112082.

10. Hong SN, Lai FTT, Wang B, Choi EPH, Wong ICK, Lam CLK, et al. Age-specific Multimorbidity Patterns and Burden on All-Cause Mortality and Public Direct Medical Expenditure: A Retrospective Cohort Study. J Epidemiol Glob Health. 2024;14(3):1077–88. 10.1007/s44197-024-00256-y.
11. Lyman GH, Carrier M, Ay C, Di Nisio M, Hicks LK, Khorana AA, et al. American Society of Hematology 2021 guidelines for management of venous thromboembolism: prevention and treatment in patients with cancer. Blood Adv. 2021;5(4):927–74. 10.1182/bloodadvances.2020003442.
12. Khorana AA, Kuderer NM, Culakova E, Lyman GH, Francis CW. Development and validation of a predictive model for chemotherapy-associated thrombosis. Blood. 2008;111(10):4902–7. 10.1182/blood-2007-10-116327.
13. Key NS, Khorana AA, Kuderer NM, Bohlke K, Lee AYY, Arcelus JI, et al. Venous Thromboembolism Prophylaxis and Treatment in Patients With Cancer: ASCO Clinical Practice Guideline Update. J Clin Oncol. 2020;38(5):496–520. 10.1200/JCO.19.01461.
14. Pallares Robles A, Ten Cate V, Lenz M, Schulz A, Prochaska JH, Rapp S, et al. Unsupervised clustering of venous thromboembolism patients by clinical features at presentation identifies novel endotypes that improve prognostic stratification. Thromb Res. 2023;227:71–81. 10.1016/j.thromres.2023.04.023.
15. Garcia-Dias R, Vieira S, Lopez Pinaya WH, Mechelli A. Clustering analysis. In: Machine Learning [Internet]. Elsevier; 2020 [cited 2026 Feb 10]. pp. 227–47. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128157398000134 10.1016/B978-0-12-815739-8.00013-4
16. Uijl A, Savarese G, Vaartjes I, Dahlström U, Brugts JJ, Linssen GCM, et al. Identification of distinct phenotypic clusters in heart failure with preserved ejection fraction. Eur J Heart Fail. 2021;23(6):973–82. 10.1002/ejhf.2169.
17. Saito Y, Omae Y, Nagashima K, Miyauchi K, Nishizaki Y, Miyazaki S, et al. Phenotyping of atrial fibrillation with cluster analysis and external validation. Heart. 2023;109(23):1751–8. 10.1136/heartjnl-2023-322447.
18. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost. 2005;3(4):692–4. 10.1111/j.1538-7836.2005.01204.x.
19. Sørensen HT, Pedersen L, Van Es N, Büller HR, Horváth-Puhó E. Impact of venous thromboembolism on the mortality in patients with cancer: a population-based cohort study. Lancet Reg Health - Eur. 2023;34:100739. 10.1016/j.lanepe.2023.100739.
20. Sørensen HT, Mellemkjær L, Olsen JH, Baron JA. Prognosis of Cancers Associated with Venous Thromboembolism. N Engl J Med. 2000;343(25):1846–50. 10.1056/NEJM200012213432504.
21. Eide NH, Langholm C, Rinde FB, Van Es N, Hveem K, Brækkan SK et al. Mortality risk after cancer-related venous thromboembolism has decreased over the last three decades: the HUNT and Tromsø studies. Haematol 2025 Feb 13. 10.3324/haematol.2024.286407

22. Crobach MJT, Anijs RJS, Brækkan SK, Severinsen MT, Hammerstrøm J, Skille H, et al. Survival after cancer-related venous thrombosis: the Scandinavian Thrombosis and Cancer Study. Blood Adv. 2023;7(15):4072–9. 10.1182/bloodadvances.2022009577.
23. Weitz JI, Haas S, Ageno W, Goldhaber SZ, Turpie AGG, Goto S, et al. Cancer associated thrombosis in everyday practice: perspectives from GARFIELD-VTE. J Thromb Thrombolysis. 2020;50(2):267–77. 10.1007/s11239-020-02180-x.
24. Investigators RIETE, Fuentes HE, Tafur AJ, Caprini JA, Alatri A, Trujillo-Santos J, et al. Prediction of early mortality in patients with cancer-associated thrombosis in the RIETE Database. Int Angiol. 2019;38(3). 10.23736/S0392-9590.19.04110-5.
25. Falanga A, Marchetti M. Cancer-associated thrombosis: enhanced awareness and pathophysiologic complexity. J Thromb Haemost. 2023;21(6):1397–408. 10.1016/j.jtha.2023.02.029.
26. Campello E, Ilich A, Simioni P, Key NS. The relationship between pancreatic cancer and hypercoagulability: a comprehensive review on epidemiological and biological issues. Br J Cancer. 2019;121(5):359–71. 10.1038/s41416-019-0510-x.
27. Poenou G, Corbaux P, Accassat S, Chauleur C, Bertoletti L. Venous Thromboembolism in Women with Breast and Gynaecological Cancers: What Do We Know and What Should We Do? Clin Appl Thromb. 2026;32:10760296251415374. 10.1177/10760296251415374.
28. Farge D, Frere C, Connors JM, Khorana AA, Kakkar A, Ay C, 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. 2022;23(7):e334–47. 10.1016/S1470-2045(22)00160-7.
29. Key NS, Khorana AA, Kuderer NM, Bohlke K, Lee AYY, Arcelus JI, et al. Venous Thromboembolism Prophylaxis and Treatment in Patients With Cancer: ASCO Guideline Update. J Clin Oncol. 2023;41(16):3063–71. 10.1200/JCO.23.00294.

Figures

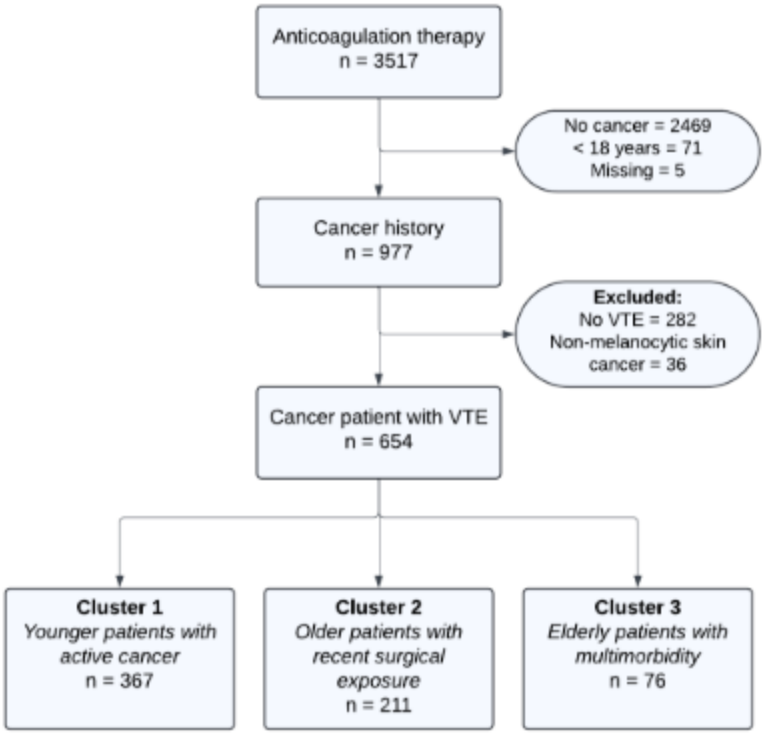


Figure 1

Flowchart of patient selection

Abbreviations: VTE: Venous thromboembolism

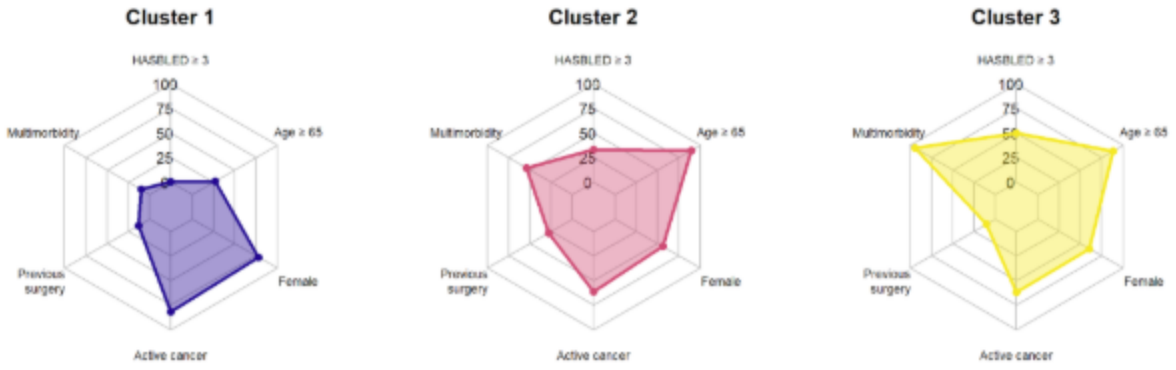


Figure 2

Radar plots showing the main clinical characteristics of each Cluster

The three clinical phenotypes were derived from unsupervised clustering based on multiple correspondence analysis followed by hierarchical clustering, as illustrated in the dendrogram provided in Supplementary Material (Figure S3). Multimorbidity was defined as the presence of two or more comorbidities.

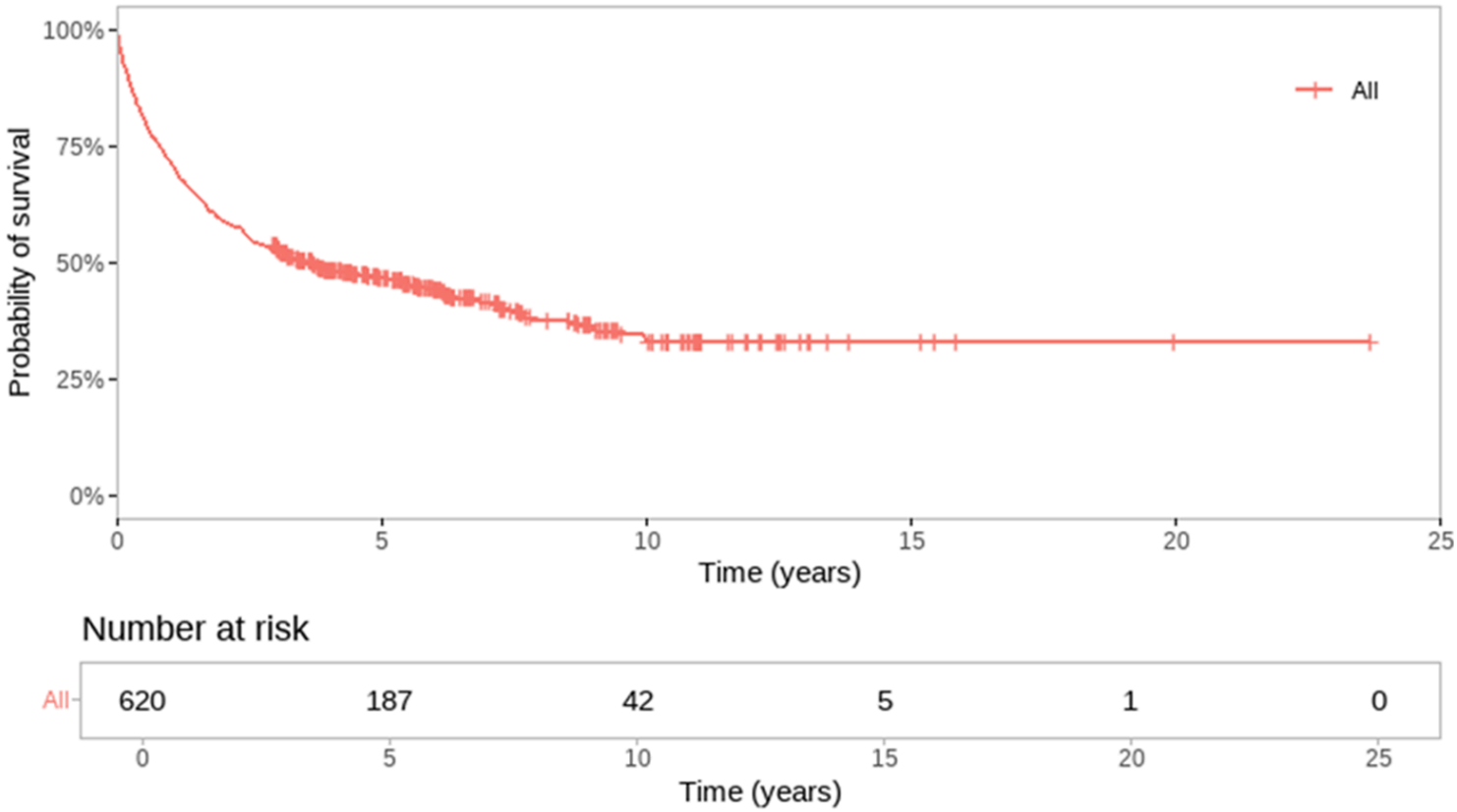


Figure 3

Overall survival (Kaplan–Meier curve)

Median overall survival was 3.7 years (95% CI: 2.8–5.4).

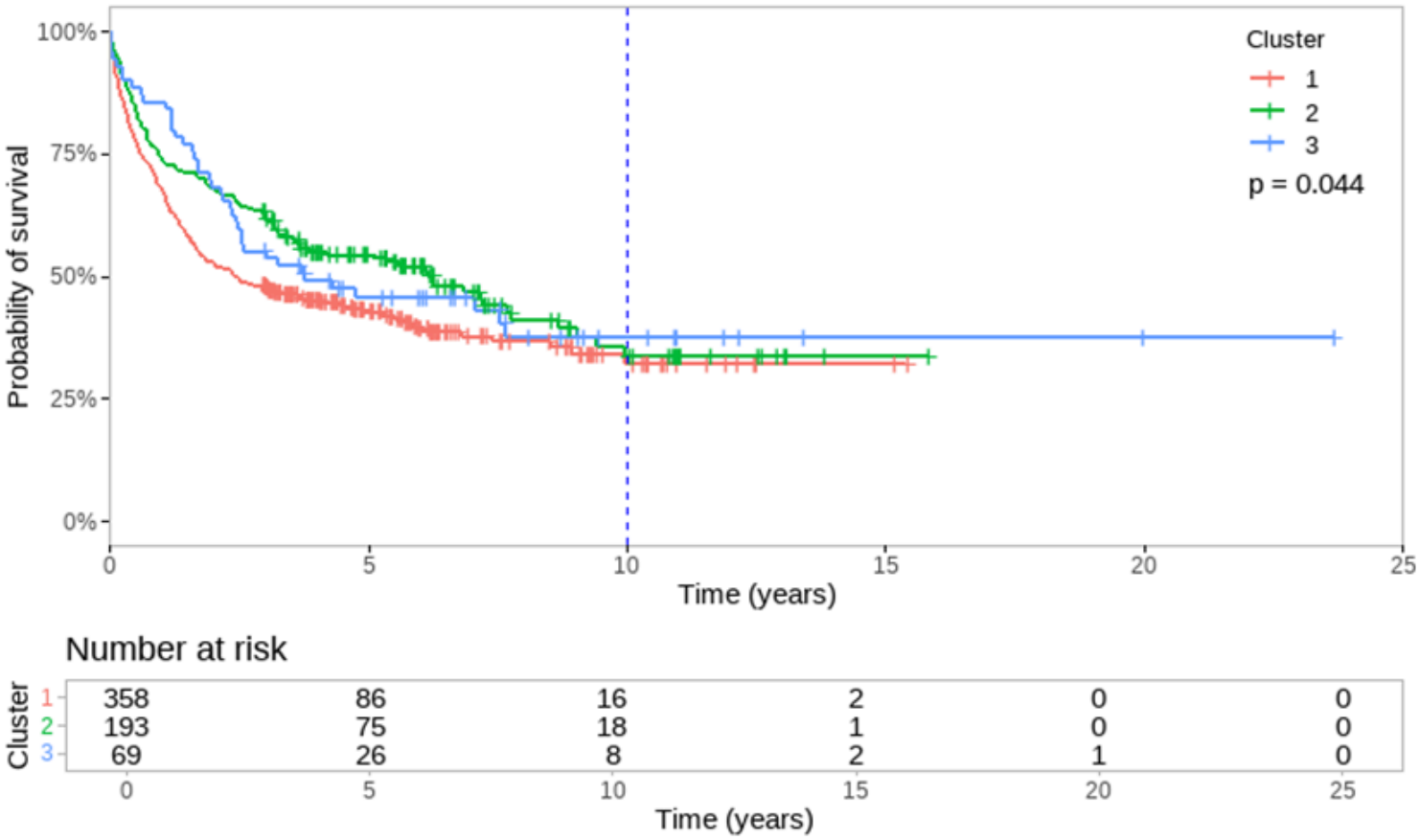


Figure 4

Survival curves stratified according to Cluster analysys (Kaplan–Meier curve)

Survival estimates beyond 10 years should be interpreted with caution due to a marked reduction in the number of patients at risk (Supplementary Material).

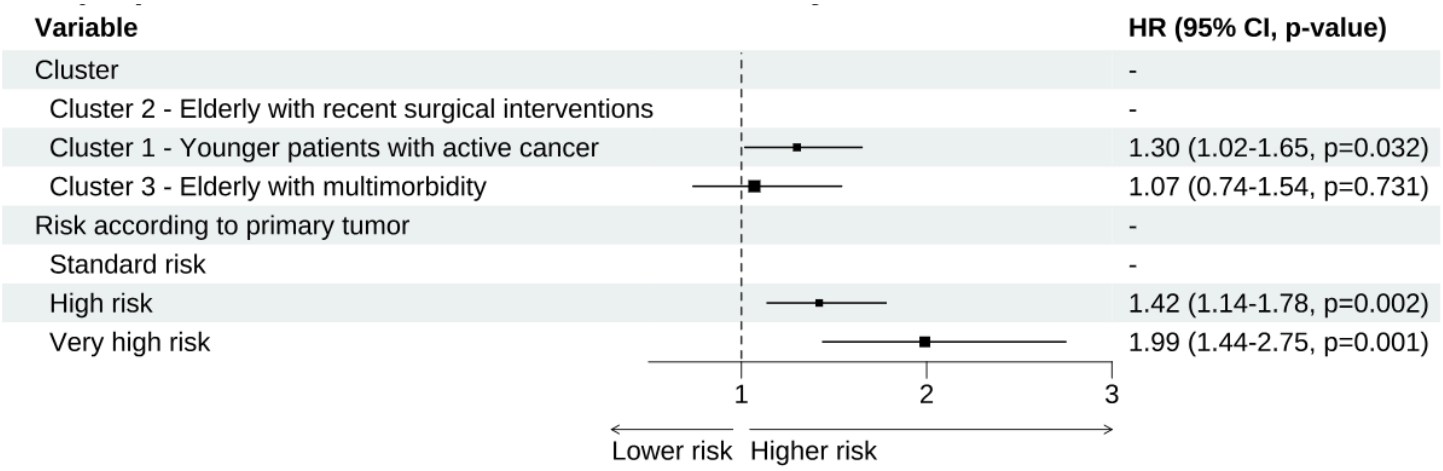


Figure 5

All-cause mortality by clinical phenotype (Multivariable Cox proportional hazards model)

Hazard ratios (HRs) and 95% confidence intervals (CIs) are shown for each clinical cluster, using Cluster 2 (older patients with recent surgical interventions) as the reference category..

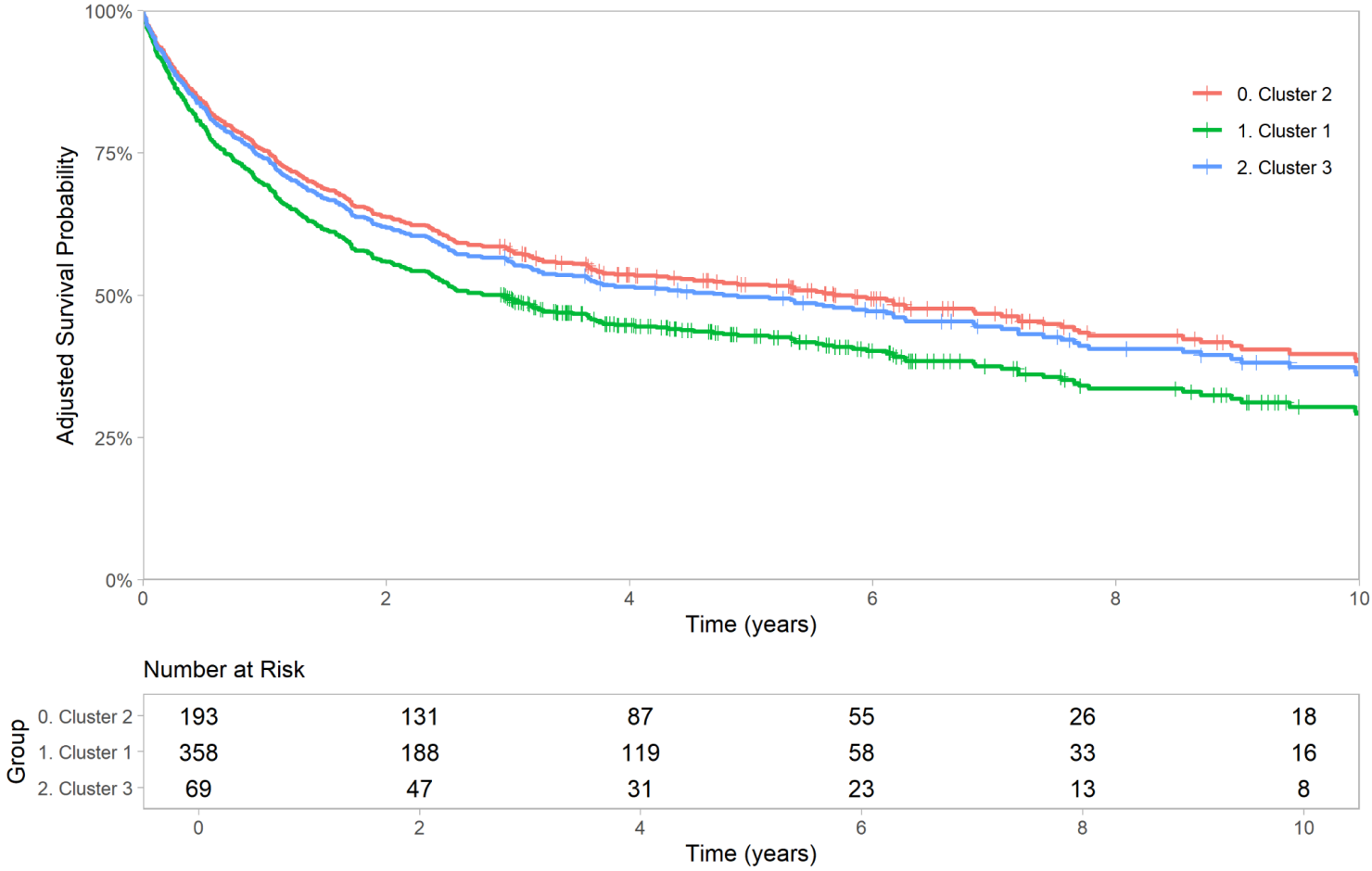


Figure 6

Adjusted Kaplan–Meier survival curves across the three clinical phenotypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx
- STROBECohort.docx