Redefining clinical venous thromboembolism phenotypes: a novel approach using latent class analysis

Phenotypic clusters among venous thrombosis patients

Maria A. de Winter *, Alicia Uijl †, Harry R. Büller ‡, Marc Carrier §, Alexander T. Cohen ¶, John-Bjarne Hansen **, Karin (H.A.H.) Kaasjager *, Lord Ajay K. Kakkar ††, Saskia Middeldorp ‡‡, Gary E. Raskob §§, Henrik Toft Sørensen ¶¶, Philip S. Wells §, Mathilde Nijkeuter * [1], Jannick A.N. Dorresteijn *** [1]

[1] Contributed equally

* Department of Acute Internal Medicine, University Medical Center Utrecht, Utrecht, the Netherlands

† Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, the Netherlands; and Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, the Netherlands; and Division of Cardiology, Department of Medicine, Karolinska Institutet, Stockholm, Sweden

‡ Department of Vascular Medicine, Amsterdam UMC, Amsterdam, the Netherlands

§ Department of Medicine, University of Ottawa, and The Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

¶ Department of Haematological Medicine, Guys and St Thomas' Hospitals, King's College London, London, United Kingdom

** Thrombosis Research Center (TREC), Department of Clinical Medicine, UiT - The Arctic University of Norway and University Hospital of North Norway, Tromsø, Norway

†† Thrombosis research Institute, London, United Kingdom

‡‡ Department of Internal Medicine & Radboud Institute of Health Sciences (RIHS), Radboud University Medical Center, Nijmegen, the Netherlands

§§ Hudson College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States

¶¶ Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

*** Department of Vascular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands

Corresponding author:
Jannick AN Dorresteijn MD, PhD
Department of Vascular Medicine, University Medical Centre Utrecht
PO Box 85500, 3508 GA Utrecht, The Netherlands
Phone: +31 88 75 555 55; Fax: +31 88 75 555 14
Email: J.A.N.Dorresteijn-2@umcutrecht.nl

Word count: 3135/5000 (manuscript), 241/250 (abstract), 3 figures, 4 tables, 1 supplementary appendix

Essentials

- Classifying venous thrombosis (VTE) patients as (un)provoked may ignore disease heterogeneity.
- Latent class analysis was used to identify phenotypic clusters among VTE patients without cancer.
- Novel clusters are young men, young women, healthy elderly, comorbidity- and history of VTE-clusters.
- Clusters differed in association with risk of recurrent VTE, bleeding and mortality.

Summary

Background Patients with venous thromboembolism (VTE) are commonly classified by presence or absence of provoking factors at time of their VTE to guide treatment decisions. This approach may not capture the heterogeneity of the disease and its prognosis.

Objectives To evaluate clinically important novel phenotypic clusters among patients with VTE without cancer, and to explore their association with anticoagulant treatment and clinical outcomes.

Patients/Methods Latent class analysis was performed with 18 baseline clinical variables in 3062 adult patients with VTE without active cancer participating in PREFER in VTE, a non-interventional disease registry. The derived latent classes were externally validated in a post-hoc analysis of Hokusai-VTE (n=6593), a randomized trial comparing edoxaban with warfarin. The associations between cluster membership and anticoagulant treatment, recurrent VTE, bleeding and mortality after initial treatment were studied.

Results Five clusters were identified: young men-cluster (n=1126, 37%), young women-cluster (n=215, 7%), older people-cluster (n=1106, 36%), comorbidity-cluster (n=447, 15%), and history of venous thromboembolism-cluster (n=168, 5%). Patient characteristics varied by age, sex, medical history, and treatment patterns. Consistent clusters were evident in external validation. In Cox proportional hazard models, recurrence risk was lower in the young women-cluster [HR 0.27 (95%CI 0.12-0.61)] compared to the comorbidity-cluster, after adjusting for extended anticoagulation. Risk of bleeding was lower in young men-, young women-, and older people-clusters [HRs 0.50 (95%CI 0.38-0.66); 0.23 (95%CI 0.11-0.46) and 0.55 (95%CI 0.41-0.73)].

Conclusions The heterogeneity of VTE cases extends beyond the distinction between provoked and unprovoked VTE.

Keywords

- Anticoagulants
- Bleeding
- Cluster analysis
- Deep venous thrombosis
- Venous thromboembolism

Introduction

Venous thromboembolism (VTE), encompassing deep venous thrombosis (DVT) and pulmonary embolism (PE), is a common condition. After the initial 3 months of treatment, anticoagulation therapy is either stopped or continued indefinitely.[1,2] To guide decisions on treatment duration, patients are commonly classified according to the presence or absence of provoking factors at the time of their VTE[3]. Until recently, a strict distinction was made, with indefinite anticoagulation constituting the treatment of choice for unprovoked VTE and 3-month treatment used for provoked VTE. However, recent studies show that risk of VTE recurrence between these two groups differs less than previously thought, weakening the rationale for this classification.[2,4,5] More recently, international guidelines recommend that anticoagulant treatment be stopped only if VTE was provoked by major transient risk factors. When no or minor provoking factors were present, and risk of bleeding is low, extended anticoagulation is suggested.[1,2]

In clinical practice, discriminating between major and minor provoking factors may be difficult. Also, this classification implies that risk of recurrent VTE follows a dichotomous or ordinal distribution. However, VTE in an individual patient is likely to be caused by many factors which accumulate until a certain threshold is exceeded and VTE occurs.[6] Thus, the simplified approach currently used may not adequately capture the disease heterogeneity of VTE, with its multifactorial etiology. This raises the question whether other approaches are needed to classify patients with VTE.

In several research areas, such as cardiology, emergency medicine, and the social sciences, phenotypic clustering has been used to improve characterization of disease phenotypes.[7–11] Applying phenotypic clustering to a population of patients with VTE

could help identifying clinically relevant patient groups and highlight currently overlooked factors. This technique may provide a starting point for improving VTE classification and further individualizing treatment decisions.

The primary aim of the present study was to identify clusters among patients with VTE. Secondary aims were to quantify the association between clusters and clinical outcomes (recurrent VTE and bleeding) and to identify possible variations in type and duration of anticoagulant treatment across different clusters.

Methods

Study populations

In this study, we used data from the PREFER in VTE registry for cluster model derivation and data from the Hokusai-VTE trial for external validation. PREFER in VTE was a non-interventional registry.[12] Between January 2013 and July 2015, consecutive adult patients with objectively confirmed symptomatic acute VTE were enrolled and followed for over 12 months. Hokusai-VTE was a randomized trial conducted between January 2010 and October 2012.[13] Adult patients with objectively confirmed acute DVT or PE were randomized to edoxaban or warfarin for 3-12 months. Exclusion criteria were contraindications to or other indications for heparin or warfarin, double antiplatelet therapy, aspirin >100 mg daily, or creatinine clearance of less than 30 ml per minute. Patients were followed for up to 12 months to assess efficacy and safety of treatment.

More detailed descriptions of these studies have been published elsewhere.[12,13] The present analysis included adult patients with VTE [i.e., DVT and/or PE] without active cancer. Active cancer was defined as active cancer with ongoing treatment noted on case report

forms (PREFER in VTE registry) or any cancer excluding non-melanoma skin cancer diagnosed or receiving cancer treatment within 6 months prior to the index VTE event, or metastatic cancer (Hokusai-VTE study). Patients with a history of cancer (i.e., any cancer diagnosis, excluding non-melanoma skin cancer, which was not classified as active cancer) were retained in the analysis. As the data collected for Hokusai-VTE were used to study associations with clinical outcomes after completion of the initial anticoagulation, the current analysis in Hokusai-VTE also excluded patients who did not complete a minimum anticoagulant treatment of 3 months.

Both studies complied with the Declaration of Helsinki. Ethical approval was obtained from the institutional review boards of the participating centers. Informed consent was provided by all patients.

Clinical variables and outcomes

Patients were characterized according to key characteristics and clinical variables considered relevant for assessing risks of recurrent VTE and bleeding. The following characteristics assessed at baseline (i.e*.,* at time of VTE diagnosis) were used: age, sex, DVT and/or PE as index event, history of cardiovascular disease, congestive heart failure, chronic lung disease, prior stroke, hypertension, history of bleeding, thrombophilia, history of VTE, diabetes mellitus (any), renal disease, history of cancer, VTE associated with estrogen therapy, VTE associated with surgery, trauma or immobilization, antiplatelet therapy, and thrombocytopenia. Clinical outcomes and anticoagulant treatment were not included in the cluster model development (unsupervised approach). The association between cluster membership and relevant clinical outcomes was assessed separately. Clinical outcomes included recurrent VTE, bleeding, and all-cause mortality after initial anticoagulant treatment.

For our analysis, patients were classified as being on extended treatment if they had received anticoagulant treatment for at least one year or if they never stopped anticoagulation. Duration of initial anticoagulant treatment was considered 90 days for patients receiving extended anticoagulation. Patients who stopped anticoagulation within one year were considered to have received initial treatment only. Definitions of clinical variables and outcomes are provided in Table S1.

Statistical analysis

Single imputation by predictive mean matching was used to impute missing values in both datasets. Novel clusters with similar clinical characteristics were identified by means of latent class analysis (LCA). LCA assumes the existence of latent classes within a population, which explain patterns of associations between clinical characteristics.[14] Each patient had a certain probability of belonging to each latent class (cluster), and is considered to belong to the cluster for which they have the highest probability. The process was repeated with different numbers of clusters (2-10). The optimal number was determined using the first minima of the Bayesian Information Criterion (BIC) and likelihood ratio test. Three models with the lowest BIC were further inspected. Among these, the option resulting in the most meaningful subgroups from a clinical perspective was chosen based on a consensus discussion between authors (JD, MN, AU, MW). As LCA is suitable only for analysis of categorical variables, continuous variables were categorized and age was analyzed per decade.[14] To explore whether the newly derived clusters corresponded to the current classification of VTE, the proportion of patients with provoked and unprovoked VTEs in each cluster was assessed. For this analysis, VTE not associated with estrogen nor with recent surgery, trauma or immobilization was considered unprovoked.

Subsequently, type and duration of anticoagulant treatment were compared among clusters in PREFER in VTE patients to assess differences in treatment patterns in clinical practice.

To study the reproducibility of our findings in external data, LCA model development was repeated in the Hokusai-VTE study. First, the cluster definitions identified in the PREFER in VTE registry were applied to the Hokusai-VTE data. Clusters and probabilities for cluster membership between the original cluster model and the cluster model derived in Hokusai-VTE were compared. The proportions of patients being classified as belonging to the same clusters are shown graphically using an alluvial plot. To assess the association between cluster membership and clinical outcomes after initial anticoagulant treatment, Cox proportional hazard models were fitted with cluster membership as a categorical covariate, using time to recurrent VTE, clinically relevant bleeding, and all-cause mortality in separate analyses as outcome variables. Combined data from the Hokusai-VTE trial and PREFER in VTE registry was used for the main analysis, to ensure sufficient numbers of outcomes per cluster. Analysis were repeated in the Hokusai-VTE trial and PREFER in VTE registry separately. Inverse probability of treatment weighting (IPTW) was used to account for possible confounding by indication. With IPTW, patients are weighted by their probability of receiving treatment using a propensity score.[15] Hazard ratios (HRs) and 95% confidence intervals (95% CIs) were reported for clusters 1, 2, 3, and 5 compared to the reference cluster 4, which was selected as the cluster with the highest event rates of recurrent VTE and bleeding. All analyses were performed using R Statistical Software, version 3.5.2. A more detailed explanation of the methodology can be found in the Supplementary Appendix.

Role of the funding source

No funding was received for the present study.

Results

Data for 3062 patients in the PREFER in VTE registry and for 6593 patients in the Hokusai-VTE trial were analyzed (Figure S1). Baseline characteristics of included patients before imputation of missing values are shown in Table 1.

Among patients enrolled in the PREFER in VTE registry, 48% were female and 56% were aged 60 years or more. In the Hokusai-VTE trial, patients were less often female (41%) and generally younger (45% aged 60 years or over). VTE was more often provoked by surgery, trauma, or immobilization in PREFER in VTE (34% versus 19%). Most comorbidities, except for chronic lung disease, and antiplatelet therapy were more frequent in PREFER in VTE.

Novel clusters

Although the 4-cluster model had the lowest BIC, the 5-cluster model was found to yield the most meaningful subgroups from a clinical perspective (*i.e.,* also including Cluster 5) (Figure S2). Baseline characteristics of patients in each cluster are shown in Table 2. Cluster 1 mostly included young men (and some women) without cardiovascular comorbidities, diabetes mellitus or cancer. Cluster 2 included younger female patients, most often with a first VTE associated with estrogen therapy. Patients in cluster 3 were the second oldest group, but comorbidity was rare, except for hypertension. Cardiovascular comorbidities as well as diabetes mellitus, chronic lung disease, and history of bleeding were most frequent in cluster 4. Compared to the other clusters, patients in cluster 4 were the oldest, most often had PE, and their VTE was most often associated with surgery, trauma, or immobilization. Most patients in cluster 5 were men and were generally somewhat older. Thrombophilia and a

history of VTE before the index event were most frequently present in this cluster, whereas VTE risk factors were not. A history of cancer was most common in clusters 3 and 5. All variables were highly distinct among clusters, except for thrombocytopenia, which was present in a minority of patients. Key characteristics of the 5 clusters are shown in Figure 1. The comorbidity-cluster and young women-cluster were the most distinct, as opposed to the history of VTE-cluster, which had the lowest median probabilities for the assigned cluster (Table S2). Probabilities of cluster membership by clinical variable are shown in Table S3 and can be used to classify any new patient according to the clusters presented.

Association with conventional classifications and treatment patterns

The distribution of patients with unprovoked VTE varied across the clusters (Table 3). In clusters 1 through 3, 53-68% had unprovoked VTE. Among patients in the history of VTE-cluster, 96% had unprovoked VTE. In the young women-cluster, in which the majority had VTE associated with estrogen use, this was only 2%. Similarly, most patients with extended treatment were in the history of VTE-cluster whereas the lowest proportion of patients with extended treatment were in the young women-cluster.

External validation

Compared to the PREFER in VTE registry, more patients in the Hokusai-VTE trial were assigned to cluster 1 (2866, 43%), whereas fewer patients could be classified into cluster 3 (1791, 27%). As patients in the Hokusai-VTE trial were generally younger and healthier, the proportion of patients aged over 60 years and the prevalence of most comorbidities was smaller in all clusters (Table 2). The distribution of relevant patient characteristics unavailable in the PREFER in VTE registry differed significantly among clusters in the Hokusai-VTE trial (Table 2). Median probabilities per cluster in the Hokusai-VTE data were comparable to those

in the PREFER in VTE data (Table S3). When repeating all steps of cluster model development in the Hokusai-VTE dataset, 5 distinct clusters, similar to the clusters found in PREFER in VTE, were identified. Again, the young women cluster was found to be most distinct (Appendix Table S4). In the Hokusai-VTE cluster model, sex seemed to be very influential, with a probability of 100% to be female in the young women-cluster, and none in the history of VTE-cluster. Age turned out to be less distinctive among clusters, as patients in the Hokusai-VTE study were generally younger. While comorbidity was less prevalent in the Hokusai-VTE study, a distinct comorbidity cluster could be identified in this study as well (Appendix Table S5). Figure 2 shows the crossover between clusters derived in the PREFER in VTE registry ('Original', left) and the Hokusai-VTE study ('Validation', right). Of the original PREFER in VTE clusters, 75% of young men, 72% of young women, 53% of patients in the older age cluster, 74% in the comorbidity cluster and 48% in the history of VTE-cluster were classified as belonging to the same cluster based on the Hokusai-VTE clusters.

Clinical outcomes

During a median follow-up of 110 days (IQR 32-228 days), 153 VTE recurrences and 125 clinically relevant bleeding events occurred in the combined population of Hokusai-VTE and PREFER in VTE registry after initial anticoagulant treatment. Crude incidence rates of recurrent VTE as well as bleeding were highest in the comorbidity-cluster and lowest in the young women-cluster (Table 4). After adjusting for the effect of extended anticoagulation during follow-up by IPTW using outcome-specific propensity scores with the comorbidity-cluster as reference group, a lower risk of recurrent VTE in patients in the young women-cluster was observed (HR 0.27; 95%CI 0.12-0.61) (Figure 3). Compared to the comorbidity-cluster, risk of bleeding in the young men-cluster, young women-cluster, and older people-cluster was significantly lower after adjustment, with respective HRs of 0.50 (95%CI 0.38-

0.66), 0.23 (95%CI 0.11-0.46), and 0.55 (95%CI 0.41-0.73). The unadjusted death rate was highest in the comorbidity-cluster and lowest in the young women-cluster. After adjusting for treatment, membership in the young women-cluster, young men-cluster, and older-people cluster was associated with a statistically significant lower risk of mortality. In PREFER in VTE registry and Hokusai-VTE, similar associations between cluster membership and outcomes were found (Appendix Tables S6 and S7, Appendix Figures S4 and S5). Contrary to the Hokusai-VTE study, in PREFER in VTE, adjusted and unadjusted risk of bleeding in the history of VTE-cluster is now found to be lower than in the comorbidity cluster.

Discussion

Five distinct, clinically relevant clusters among the population of patients with VTE (without cancer) were identified in data from a prospective cohort study and validated in a large randomized trial. These were a young men-cluster, a young women-cluster, an older age-cluster, a comorbidity-cluster, and a history of VTE-cluster. Patients within each cluster differed considerably in terms of clinical characteristics, treatment patterns, and associations with recurrent VTE, clinically relevant bleeding, and mortality. This emphasizes the heterogeneity among patients with VTE.

The novel clusters partly overlap with the current classification based on presence of provoking factors.[2] Most patients in the history of VTE-cluster had unprovoked VTE, whereas patients in the young women-cluster had VTE associated with estrogen therapy. In line with current guidelines, anticoagulant treatment duration in these clusters is reflective of the distribution of provoking factors, with the highest proportion of patients with extended treatment in the history of VTE-cluster and the lowest proportion of patients with extended

treatment in the young women-cluster.[1,16] Traditional provoking factors do not seem to distinctly define the other 3 clusters, nor does the proportion of patients with extended treatment vary much among those clusters.

Our novel classification highlights the importance of age and comorbidities when characterizing patients with VTE. Two clusters consist mainly of younger patients, whereas patients in the older people-cluster, the comorbidity-cluster, and the history of VTE-cluster are generally older. Which of the 3 clusters best describes an individual older patient is determined by comorbid conditions and absence of provoking factors. Age and comorbidity are also very important in explaining differences in associations with clinical outcomes. Risks of bleeding and death increase with older age, as does the prevalence of several risk factors for recurrent VTE.[1,2,17]

In the external validation cluster model, derived in patients with less variation in age, sex was found to be important too when characterizing patients. This explains the large proportion of patients originally classified as belonging to the older people-cluster who are reclassified to the history of VTE-cluster in the Hokusai-VTE clustering model (Figure 2). The older people cluster includes a mixed population in the original cluster model, whereas in the Hokusai-VTE cluster model, this cluster includes hardly only females. The history of VTE-cluster, on the other hand, includes only male patients in the Hokusai-VTE model. Potential explanations could be that, in Hokusai-VTE, older men more often have comorbidities (or a history of VTE), and/or that older women participating in Hokusai-VTE are generally healthier than older men. Alternatively, this may reflect that causes underlying VTE may be different in men and women.

The young men-cluster and young women-cluster both include relatively young, generally healthy patients. Roughly speaking, the young men-cluster includes young men and young women not using estrogen therapy. Female-specific risk factors are often used to characterize women with VTE.[18] Risk of first-time VTE is generally higher for men than for women, except for women of reproductive age. Compared to the young women-cluster, members of the young men-cluster are somewhat older and comorbidities are more common. Risks of recurrent VTE, bleeding and mortality were lowest in the young women-cluster, in line with previous literature.[19,20] Although premenopausal women may be at risk of heavy menstrual bleeding, risk of bleeding was lowest in this cluster. A potential explanation could be that abnormal uterine bleeding is not always adequately measured in studies. Other studies have shown that risk of recurrence is similar in patients with and without hormone therapy when restricted to young women.[20,21] These findings are likely reflective of the young population and its absence of comorbidities. When looking at recurrent VTE, available literature shows that the risk is low if estrogen therapy is discontinued when anticoagulation is stopped.[22,23]

Risks of recurrent VTE, bleeding, and mortality were highest in the comorbidity-cluster, which had the oldest patients and the highest prevalence of cardiovascular and other comorbidities. This emphasizes the importance of individualized decision-making for patients in this cluster. The role of cardiovascular comorbidity in characterizing this cluster is unsurprising. Previous studies have shown that obesity, diabetes mellitus, hypertension, low HDL-cholesterol levels, and high triglyceride levels are not only associated with older age and arterial cardiovascular disease, but also with VTE.[24] Coexistence of multiple major cardiovascular risk factors has an additive causative effect and contributes to the multifactorial nature of VTE.[24] Recognizing the association between cardiovascular risk factors and VTE suggests new

preventive strategies, which may include weight loss, lipid-lowering therapy, and a more prominent role for antiplatelet therapy. In selected patients, antiplatelet therapy already has a place in the secondary prevention of VTE.[1,2] Furthermore, statins have been shown to have a modest protective effect on both first-time and recurrent VTE.[25,26] Our findings underscore the importance of cardiovascular risk assessment in patients with VTE.

In the older people-cluster, comorbidities other than hypertension were uncommon. This may explain why risks of recurrent VTE and bleeding in this cluster were not significantly higher than in the young men-group. The lower risk of recurrent VTE and bleeding in elderly patients without comorbidity compared to elderly, multimorbid patients has been described previously.[27] Members of the history of VTE-cluster have comorbidity more frequently. More importantly, these patients have the highest prevalence of well-known risk factors for recurrent VTE, as well as bleeding. These include the highest prevalence of unprovoked VTE, history of VTE before the index event, the highest BMI, and also the lowest eGFR and platelet count, highest blood pressure values, and highest prevalence of alcohol use. As cardiovascular disease is common in the history of VTE-cluster, many patients are on antiplatelet therapy.[28] This explains why risk of bleeding in the history of VTE-cluster is comparable to that in the comorbidity-cluster, even after adjusting for anticoagulant treatment. Surprisingly, risk of recurrent VTE was not found to be elevated. In PREFER in VTE registry, risk of bleeding in the history of VTE-cluster was lower compared to the comorbidity cluster. A potential explanation could be that the difference in prevalence of antiplatelet therapy use between the clusters is larger in this study.

Our study is the first to use cluster analysis to characterize patients with VTE without active cancer based on clinical characteristics. Comparable studies have used cluster analysis only

in the context of antiphospholipid syndrome, D-dimer levels in COVID-19, and imaging parameters and risk of hemodynamic deterioration among patients with PE, but never in the total population of patients with VTE after the initial anticoagulant treatment.[29–31] We identified distinct and clinically plausible clusters and our findings were confirmed when repeating the analyses in external data. However, some important limitations need to be addressed. First, some variables which were considered relevant a priori were not available in the PREFER in VTE dataset. Therefore, pregnancy or puerperium, chronic inflammatory disease, use of NSAIDs, overweight or obesity, and anemia could not be used to define clusters. Second, our findings apply only to patients who resemble the study population of PREFER in VTE, from which the clusters were derived. However, as an international prospective cohort with few exclusion criteria, the PREFER in VTE study population is likely to be representative of the total population. Additionally, findings were confirmed in a randomized trial, which may represent a different subset of the population. However, patients with extremes of age may have been underrepresented. Third, studying outcomes in combined data of patients with and without extended anticoagulation may cause confounding by indication. However, this has been adjusted for using IPTW. Finally, of note, cluster analysis is a data-driven technique that aims to generate hypotheses for future research. Although clinically plausible, the identified clusters represent statistical associations and do not necessarily reflect pathophysiology. Hence, our results should be starting points for future research rather than strict recommendations for clinical practice. For example, the results could be used when designing trials, *e.g*., to plan subgroup analyses to assess treatment effect heterogeneity among clusters, or to ensure that all clusters are represented in a trial's study population. These approaches may help to formulate recommendations for clinically relevant subgroups.

In conclusion, among the population of patients with VTE without active cancer, five distinct phenotypic clusters were identified that differed by clinical characteristics and prognosis. These findings emphasize the heterogeneity of VTE cases, which extends beyond the simple distinction between provoked and unprovoked VTE. Differences in risk of bleeding and recurrent VTE between clusters suggest that cluster-specific treatment recommendations need to be explored to improve clinical decision-making on duration of treatment.

Addendum

M.A. de Winter, A. Uijl, M. Nijkeuter, and J.A.N. Dorresteijn were responsible for study conception and design. M.A. de Winter performed the data analysis and drafted the manuscript in close collaboration with A. Uijl, J.A.N. Dorresteijn, and M. Nijkeuter. All authors contributed to interpretation of the results, critically revised the manuscript, and were responsible for and approved its final version.

Acknowledgements

This publication is based on research using data from Daiichi Sankyo, Inc. that has been made available through Vivli, Inc. Vivli has not contributed to or approved, and is not in any way responsible for the contents of this publication.

Conflicts of interest

M. Carrier reports research funding from BMS, Leo Pharma and Pfizer, and honorarium from Bayer, Sanofi, BMS, Leo Pharma, Servier and Pfizer. All fees are paid to his institution. A.T. Cohen reports consulting fees, advisory boards and research support from Abbott, AbbVie,

ACI Clinical, Alexion Pharmaceuticals, Aplagon Ltd, Bayer, Boehringer-Ingelheim, Boston Scientific, Bristol-Myers Squibb, BTG, Cambridge Healthcare Research, Daiichi-Sankyo, EmstoPA Ltd, EPG Health, Guidepoint Global, Gulf Coast Developments, Janssen, Johnson & Johnson, Leo Pharma, LifeSciences Consulting, Medscape, McKinsey, Navigant, North Star Communications, ONO Pharmaceuticals, Pfizer, Portola Pharmaceuticals, Sanofi, Takeda, Total CME and Windrose Consulting Group. A.K. Kakkar reports research support from Bayer Sanofi and personal fees from Bayer Sanofi and Anthos Inc. S. Middeldorp reports grants and personal fees from Daiichy Sankyo, grants and personal fees from Bayer, grants and personal fees from Pfizer, grants and personal fees from Boehringer-Ingelheim, personal fees from Portola/Alexion, personal fees from Abbvie, personal fees from BMS Pfizer, personal fees from Norgine. All fees are paid to her institution. G.E. Raskob reports consultancy fees or honoraria from Bayer HealthCare Pharmaceuticals Inc., Bristol-Myers Squibb, Daiichi Sankyo Inc., Eli Lilly and Company, Itreas, Janssen Global Services LLC, Medscape, Novartis, Pfizer, Portola Pharmaceuticals, Teherex and XaTrek; honoraria from ACC Oklahoma; Canadian Cardiovascular Society, Japanese Circulation Society, M3 LTD, MD Magazine and North American Thrombosis Forum; stock or stock option ownership for AbbVie, Inc., ACADIA Pharmaceuticals Inc, Altimmune Inc, Arch Therapeutics, Atea Pharmaceuticals, Biomarin, Cohbar, Eli Lilly and Company, EsperionTherapeutics, Gilead Sciences INc, GlaxoSmithKline,LLC., Incyte Corporation, Merck, Moderna, NKTR, Pfizer, Precigen, Sangamo Therapeutics Inc, Vaxart and ZioPharm Oncology. P.S. Wells reports grants and speaker fees from BMS/Pfizer, and speaker fees from Bayer Healthcare. All fees are paid to his institution. The other authors have nothing to declare.

References

1 Kearon C, Akl EA, Ornelas J, Blaivas A, Jimenez D, Bounameaux H, Huisman M, King CS, Morris TA, Sood N, Stevens SM, Vintch JRE, Wells P, Woller SC, Moores L. Antithrombotic Therapy for VTE Disease. *Chest* 2016; **149**: 315–52.

2 Konstantinides S V, Meyer G, Becattini C, Bueno H, Geersing G-J, Harjola V-P, Huisman M V, Humbert M, Jennings CS, Jiménez D, Kucher N, Lang IM, Lankeit M, Lorusso R, Mazzolai L, Meneveau N, Ní Áinle F, Prandoni P, Pruszczyk P, Righini M, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). *Eur Heart J* England; 2019; : ehz405.

3 Kearon C, Ageno W, Cannegieter SC, Cosmi B, Geersing G-J, Kyrle PA. Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH. *J Thromb Haemost* England; 2016; **14**: 1480–3.

4 Albertsen IE, Nielsen PB, Søgaard M, Goldhaber SZ, Overvad TF, Rasmussen LH, Larsen TB. Risk of Recurrent Venous Thromboembolism: A Danish Nationwide Cohort Study. *Am J Med* Elsevier Inc.; 2018; **131**: 1067-1074.e4.

5 Ageno W, Farjat A, Haas S, Weitz JI, Goldhaber SZ, Turpie AGG, Goto S, Angchaisuksiri P, Dalsgaard Nielsen J, Kayani G, Schellong S, Bounameaux H, Mantovani LG, Prandoni P, Kakkar AK. Provoked versus unprovoked venous thromboembolism: Findings from GARFIELD-VTE. *Res Pract Thromb Haemost* 2021; **5**: 326–41.

6 Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet (London, England)* England; 1999; **353**: 1167–73.

7 Ahmad T, Pencina MJ, Schulte PJ, O'Brien E, Whellan DJ, Piña IL, Kitzman DW, Lee KL, O'Connor CM, Felker GM. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol* 2014; **64**: 1765–74.

8 Seymour CW, Kennedy JN, Wang S, Chang CCH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H, Huang DT, Kellum JA, Mi Q, Opal SM, Talisa V, Van Der Poll T, Visweswaran S, Vodovotz Y, Weiss JC, Yealy DM, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA - J Am Med Assoc* 2019; **321**: 2003–17.

9 Geiser C, Lehmann W, Eid M. Separating "Rotators" From "Nonrotators" in the Mental Rotations Test: A Multigroup Latent Class Analysis. *Multivariate Behav Res* United States; 2006; **41**: 261–93.

10 Inohara T, Shrader P, Pieper K, Blanco RG, Thomas L, Singer DE, Freeman J V., Allen LA, Fonarow GC, Gersh B, Ezekowitz MD, Kowey PR, Reiffel JA, Naccarelli G V., Chan PS, Steinberg BA, Peterson ED, Piccini JP. Association of of atrial fibrillation clinical phenotypes with treatment patterns and outcomes a multicenter registry study. *JAMA Cardiol* 2018; **3**: 54–63.

11 Uijl A, Savarese G, Vaartjes I, Dahlström U, Brugts JJ, Linssen GC, Empel V, Rocca HB, Asselbergs FW, Lund LH, Hoes AW, Koudstaal S. Identification of Distinct Phenotypic Clusters in Heart Failure with Preserved Ejection Fraction. *Eur J Heart Fail* 2021; **23**: 973–82.

12 Cohen AT, Gitt AK, Bauersachs R, Fronk EM, Laeis P, Mismetti P, Monreal M, Willich SN, Bramlage P, Agnelli G. The management of acute venous thromboembolism in clinical practice results from the european prefer in vte registry. *Thromb Haemost* 2017; **117**: 1326–37.

13 Büller HR, Décousus H, Grosso MA, Mercuri M, Middeldorp S, Prins MH, Raskob GE, Schellong SM, Schwocho L, Segers A, Shi M, Verhamme P, Wells P. Edoxaban versus warfarin for the treatment of symptomatic venous thromboembolism. *N Engl J Med* 2013; **369**: 1406–15.
14 Linzer DA, Lewis JB. poLCA: An R package for polytomous variable latent class analysis. *J Stat Softw* 2011; **42**: 1–29.
15 Mitchell JD, Gage BF, Fergestrom N, Novak E, Villines TC. Inverse Probability of Treatment Weighting (Propensity Score) using the Military Health System Data Repository and National Death Index. *J Vis Exp* 2020; .
16 Konstantinides S V, Torbicki A, Agnelli G, Danchin N, Fitzmaurice D, Galiè N, Gibbs JSR, Huisman M V, Humbert M, Kucher N, Lang I, Lankeit M, Lekakis J, Maack C, Mayer E, Meneveau N, Perrier A, Pruszczyk P, Rasmussen LH, Schindler TH, et al. 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J* England; 2014; **35**: 3033-3069k.
17 Tritschler T, Castellucci LA, van Es N, Aujesky D, Le Gal G. Treatment of venous thromboembolism in elderly patients in the era of direct oral anticoagulants. *Polish Arch Intern Med* 2020; **130**: 529–38.
18 Roach REJ, Cannegieter SC, Lijfering WM. Differential risks in men and women for first and recurrent venous thrombosis: The role of genes and environment. *J Thromb Haemost* 2014; **12**: 1593–600.
19 Eischer L, Eichinger S, Kyrle PA. The risk of recurrence in women with venous thromboembolism while using estrogens: A prospective cohort study. *J Thromb Haemost* 2014; **12**: 635–40.
20 Skeith L, Le Gal G, Rodger MA. Oral contraceptives and hormone replacement therapy: How strong a risk factor for venous thromboembolism? *Thromb Res* 2021; **202**: 134–8.
21 Kyrle PA, Minar E, Bialonczyk C, Hirschl M, Weltermann A, Eichinger S. The Risk of Recurrent Venous Thromboembolism in Men and Women. *N Engl J Med* Massachusetts Medical Society; 2004; **350**: 2558–63.
22 Douketis J, Tosetto A, Marcucci M, Baglin T, Cosmi B, Cushman M, Kyrle P, Poli D, Tait RC, Iorio A. Risk of recurrence after venous thromboembolism in men and women: Patient level meta-analysis. *Bmj* 2011; **342**: 535.
23 Kiconco S, Abdul Sultan A, Grainge MJ. Recurrence risk of venous thromboembolism and hormone use in women from England: a cohort study using clinical practice research datalink. *Br J Haematol* England; 2017; **177**: 127–35.
24 Ageno W, Becattini C, Brighton T, Selby R, Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism: A meta-analysis. *Circulation* 2008; **117**: 93–102.
25 Schmidt M, Cannegieter SC, Johannesdottir SA, Dekkers OM, Horváth-Puhó E, Sørensen HT. Statin use and venous thromboembolism recurrence: A combined nationwide cohort and nested case-control study. *J Thromb Haemost* 2014; **12**: 1207–15.
26 Ray JG, Rosendaal FR. The role of dyslipidemia and statins in venous thromboembolism. *Curr Control Trials Cardiovasc Med* Department of Medicine, University of Toronto, Toronto, and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. rayjg@mcmaster; 2001; **2**: 165–70.
27 Lange N, Méan M, Stalder O, Limacher A, Tritschler T, Rodondi N, Aujesky D. Anticoagulation quality and clinical outcomes in multimorbid elderly patients with

acute venous thromboembolism. *Thromb Res* Elsevier; 2019; **177**: 10–6.
28 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* Elsevier Ltd; 2021; **199**: 85–96.
29 Zuily S, Clerc-Urmès I, Bauman C, Andrade D, Sciascia S, Pengo V, Tektonidou MG, Ugarte A, Gerosa M, Michael Belmont H, Zamorano MAA, Fortin P, Ji L, Efthymiou M, Cohen H, Branch DW, Jesus GR de, Nalli C, Petri M, Rodriguez E, et al. Cluster analysis for the identification of clinical phenotypes among antiphospholipid antibody-positive patients from the APS ACTION Registry. *Lupus* 2020; .
30 Naymagon L, Zubizarreta N, Feld J, Gerwen M Van, Alsen M, Thibaud S, Kessler A, Venugopal S, Makki I, Qin Q, Dharmapuri S, Jun T, Bhalla S, Berwick S, Christian K. Admission D-dimer levels, D-dimer trends, and outcomes in COVID-19 Leonard. *Thromb Res* 2020; **196**: 99–105.
31 Gao Y, Wang N, Li Y, Huang H, Jia D. Phenotyping of non-high-risk acute pulmonary embolism patients: different initial manifestations of clinical deterioration. *Eur J Intern Med* 2021; **90**: 116–8.

Tables

Table 1. Baseline characteristics of patients enrolled in the PREFER in VTE registry and in the Hokusai-VTE trial.

	PREFER in VTE		Hokusai-VTE	
	n (% of 3062)	n (%) missing [a]	n (% of 6593)	n (%) missing [a]
Female sex	1381 (47.3)	0 (0.0)	2706 (41.0)	0 (0.0)
Age >60 years	1620 (55.5)	0 (0.0)	2933 (44.5)	0 (0.0)
PE as index event	1201 (41.2)	0 (0.0)	2431 (36.9)	0 (0.0)
Provoking factors				
Estrogen therapy	213 (7.3)	1 (0.0)	479 (7.3)	9 (0.1)
Surgery, trauma, or immobilization	968 (33.2)	4 (0.1)	1245 (18.9)	9 (0.1)
Medical history				
Prior stroke	135 (4.4)	3 (0.1)	193 (2.9)	0 (0.0)
History of bleeding	107 (3.5)	0 (0.0)	51 (0.8)	0 (0.0)
Hypertension	1298 (42.2)	2 (0.1)	2554 (38.7)	0 (0.0)
Cardiovascular disease	287 (9.8)	7 (0.2)	558 (8.5)	0 (0.0)
Congestive heart failure	120 (3.9)	3 (0.1)	175 (2.7)	0 (0.0)
Known thrombophilia	176 (5.7)	0 (0.0)	371 (5.6)	9 (0.1)
History of VTE prior to index event	748 (24.4)	1 (0.0)	1247 (18.9)	0 (0.0)
Chronic lung disease	267 (8.7)	2 (0.1)	741 (11.2)	0 (0.0)
Diabetes mellitus	306 (10.0)	2 (0.1)	664 (10.1)	2 (0.0)
Renal disease	174 (5.9)	0 (0.0)	96 (1.5)	5 (0.1)
History of cancer	250 (8.2)	0 (0.0)	405 (6.1)	0 (0.0)
Medication use				
Antiplatelet therapy	346 (11.4)	16 (0.5)	1341 (20.3)	0 (0.0)
Laboratory measurements				
Thrombocytopenia	12 (0.4)	4 (0.1)	4 (0.1)	35 (0.5)

Abbreviations: IQR, interquartile range; PE, pulmonary embolism; VTE, venous thromboembolism

a Values for one or more clustering variables were missing in 37 (1%) of patients in the PREFER in VTE registry and 46 (1%) of patients in the Hokusai-VTE trial.

Table 2. Patient characteristics stratified by cluster in the PREFER in VTE registry and the Hokusai-VTE trial.
All data are shown as n (% of total per cluster) or median (interquartile range) per cluster. Cluster 1: young men; cluster 2: young women, cluster 3: older age; cluster 4: comorbidity; cluster 5: history of VTE. *Probabilities are based on the original cluster model, developed in the PREFER in VTE registry.*

	PREFER in VTE (n=3062)						**Hokusai-VTE** (n=6593)					
	1	**2**	**3**	**4**	**5**	**p**	**1**	**2**	**3**	**4**	**5**	**p**
N	1126 (37)	215 (7)	1106 (36)	447 (15)	168 (5)	n.a.	2866 (43.5)	420 (6.4)	1791 (27.2)	1058 (16.0)	458 (7.0)	n.a.
Female sex [a], [b, c]	392 (34.8)	213 (99.1)	576 (52.1)	219 (49)	62 (36.9)	<0.001	810 (28.3)	419 (99.8)	886 (49.5)	450 (42.5)	141 (30.8)	<0.001
Age >60 years [a], [b, c]	220 (19.5)	4 (1.9)	946 (85.5)	404 (90.4)	145 (86.3)	<0.001	428 (14.9)	8 (1.9)	1354 (75.6)	835 (78.9)	308 (67.2)	<0.001
PE as index event [a], [b]	357 (31.7)	88 (40.9)	534 (48.3)	215 (48.1)	49 (29.2)	<0.001	845 (29.5)	206 (49.0)	750 (41.9)	499 (47.2)	131 (28.6)	<0.001
Provoking factors												
Estrogen therapy [a], [b]	0 (0.0)	207 (96.3)	8 (0.7)	1 (0.2)	1 (0.6)	<0.001	2 (0.1)	405 (96.4)	20 (4.4)	15 (1.4)	38 (2.1)	<0.001
Surgery, trauma, or immobilization [a], [b]	388 (34.5)	76 (35.3)	347 (31.4)	211 (47.2)	5 (3.0)	<0.001	596 (20.8)	72 (17.1)	339 (18.9)	231 (21.8)	9 (2.0)	<0.001
Pregnancy [a], [b]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	15 (0.5)	1 (0.2)	0 (0.0)	2 (0.2)	0 (0.0)	0.005
Medical history												
Prior stroke [b]	0 (0.0)	3 (1.4)	9 (0.8)	123 (27.5)	0 (0.0)	<0.001	0 (0.0)	1 (0.2)	27 (1.5)	165 (15.5)	0 (0.0)	<0.001
History of bleeding [b]	30 (2.7)	4 (1.9)	35 (3.2)	38 (8.5)	0 (0.0)	<0.001	19 (0.7)	1 (0.2)	8 (0.4)	23 (2.2)	0 (0.0)	<0.001
Hypertension [b]	59 (5.2)	14 (6.5)	745 (67.4)	355 (79.4)	126 (75)	<0.001	184 (6.4)	38 (9.0)	1226 (68.5)	821 (77.6)	285 (62.2)	<0.001
Cardiovascular disease [b]	12 (1.1)	0 (0.0)	5 (0.5)	271 (60.6)	23 (13.7)	<0.001	23 (0.8)	0 (0.0)	21 (1.2)	473 (44.7)	41 (9.0)	<0.001

Congestive heart failure [a], [b]	0 (0.0)	0 (0.0)	28 (2.5)	92 (20.6)	0 (0.0)	<0.001	0 (0.0)	1 (0.2)	45 (2.5)	129 (12.2)	0 (0.0)	<0.001
Known thrombophilia [a]	86 (7.6)	18 (8.4)	0 (0.0)	12 (2.7)	60 (35.7)	<0.001	215 (7.5)	18 (4.3)	0 (0.0)	15 (1.4)	125 (27.3)	<0.001
Previous VTE before index event [a], [b]	317 (28.2)	14 (6.5)	157 (14.2)	109 (24.4)	151 (89.9)	<0.001	540 (18.5)	16 (3.8)	226 (12.6)	159 (15.0)	316 (69.0)	<0.001
Chronic lung disease [a], [b]	40 (3.6)	4 (1.9)	121 (10.9)	96 (21.5)	6 (3.6)	<0.001	154 (5.4)	33 (7.9)	271 (15.1)	250 (23.6)	33 (7.2)	<0.001
Diabetes mellitus [b], [c]	4 (0.4)	0 (0.0)	162 (14.6)	109 (24.4)	32 (19)	<0.001	38 (1.3)	0 (0.0)	309 (17.3)	245 (23.2)	73 (15.9)	<0.001
Renal disease [c]	0 (0.0)	1 (0.5)	82 (7.4)	73 (16.3)	18 (10.7)	<0.001	0 (0.0)	0 (0.0)	30 (1.7)	54 (5.1)	10 (2.2)	<0.001
History of cancer [a], [b], [c]	11 (1.0)	6 (2.8)	158 (14.3)	51 (11.4)	24 (14.3)	<0.001	26 (0.9)	10 (2.4)	212 (11.8)	102 (9.6)	55 (12.0)	<0.001
Chronic inflammatory disease	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	57 (2.0)	6 (1.4)	75 (4.2)	48 (4.5)	19 (4.1)	<0.001
Post thrombotic syndrome	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	23 (0.8)	0 (0.0)	8 (0.4)	8 (0.8)	12 (2.6)	<0.001
Medication use												
Antiplatelet therapy [b]	3 (0.3)	0 (0.0)	0 (0.0)	287 (64.2)	56 (33.3)	<0.001	209 (7.3)	8 (1.9)	14 (0.8)	840 (79.4)	270 (59.0)	<0.001
Estrogen therapy	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	71 (2.5)	80 (19.0)	10 (0.6)	5 (0.5)	8 (1.7)	<0.001
NSAIDs	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	742 (25.9)	134 (31.9)	421 (23.5)	289 (27.3)	122 (26.6)	0.005
Statins	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	180 (6.3)	18 (4.3)	334 (18.6)	480 (45.4)	139 (30.3)	<0.001
Corticosteroids	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	226 (7.9)	45 (10.7)	269 (15.0)	213 (20.2)	55 (12.0)	<0.001
Laboratory measurements												

Thrombocytopenia	2 (0.2)	0 (0.0)	8 (0.7)	2 (0.4)	0 (0.0)	0.200	1 (0.0)	0 (0.0)	1 (0.1)	2 (0.0)	0 (0.2)	0.444
Platelet count [b]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	222 (186-262)	248 (212-288)	216 (183-256)	213 (179-258)	212 (178-251)	<0.001
BMI (kg/m^2) [b]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	27.2 (24.4-30.7)	26.9 (23.0-31.2)	28.1 (25.1-31.9)	28.7 (25.4-32.4)	29.1 (26.1-33.3)	<0.001
eGFR (mL/min) [b]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	79.9 (65.9-103.2)	116.5 (107.1-126.9)	73.7 (56.7-91.0)	68.4 (51.7-87.1)	67.8 (53.7-87.0)	<0.001
Hemoglobin (g/dL) [b]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	14.5 (13.3-15.4)	13.4 (12.6-13.9)	13.8 (12.8-14.8)	13.7 (12.6-14.7)	14.2 (13.2-15.1)	<0.001
Other												
Systolic blood pressure (mmHg) [b]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	125 (120-131)	120 (110-128)	130 (120-140)	130 (120-140)	130 (120-140)	<0.001
Smoking status (yes)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	779 (27.2)	91 (21.7)	217 (12.1)	162 (15.3)	91 (19.9)	<0.001
Alcohol use	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	1087 (37.9)	157 (37.4)	569 (31.3)	307 (29.0)	200 (43.7)	<0.001

Abbreviations: BMI, body mass index; eGFR, estimated glomerular filtration rate; IQR, interquartile range; PE, pulmonary embolism; NSAIDs, non-steroidal anti-inflammatory drugs; VTE, venous thromboembolism

a Variables included to estimate the propensity score for recurrent VTE.

b Variables included to estimate the propensity score for bleeding. For continuous variables unavailable in one of the datasets, the mean value of the other dataset was imputed.

c Variables included to estimate the propensity score for mortality.

Table 3. Association with conventional classification and treatment patterns among patients in the PREFER in VTE registry, stratified by cluster.
All data are shown as n (% of total per cluster).

	Young men	**Young women**	**Older people**	**Comorbidity**	**History of VTE**	**p-value**
n	1126	215	1106	447	168	n.a.
Unprovoked VTE (%)	738 (70)	7 (3)	751 (68)	235 (53)	162 (96)	<0.001
n (%) with data available on extended treatment	912 (81)	189 (88)	923 (83)	341 (76)	151 (90)	n.a.
Median (IQR) duration of anticoagulant treatment in days	211 (117-365)	184 (97-360)	344 (160-366)	359 (172-367)	363 (208-371)	<0.001
Extended treatment in days	363 (39)	52 (27)	463 (50)	190 (56)	106 (70)	<0.001
Type of extended treatment - VKA - DOAC - Other	 245 (68) 91 (25) 27 (7)	 34 (65) 16 (31) 2 (4)	 324 (70) 102 (22) 37 (8)	 147 (78) 26 (14) 17 (9)	 78 (74) 24 (23) 4 (4)	0.530

Abbreviations: DOAC, direct oral anticoagulant; IQR, interquartile range; VKA, vitamin K antagonist; VTE, venous thromboembolism

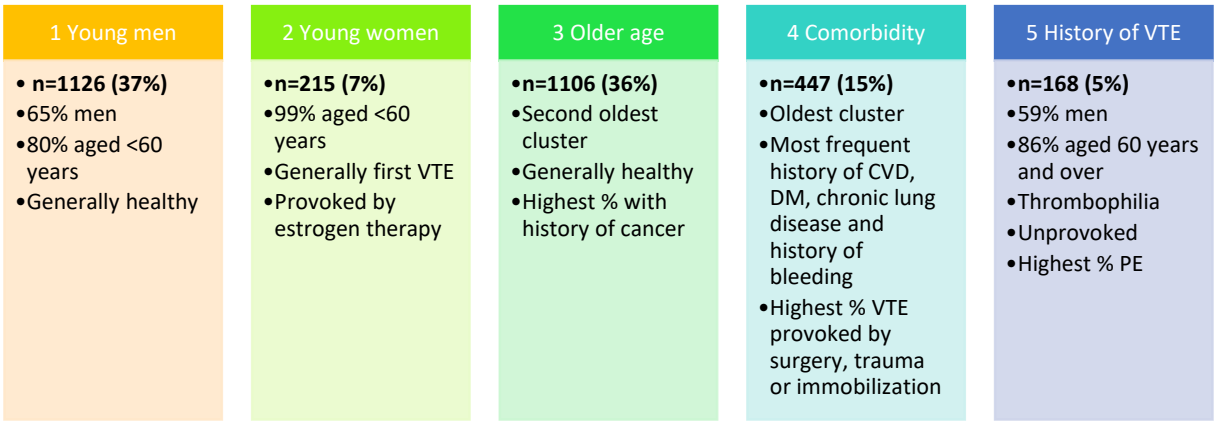
Table 4. Association between cluster membership and clinical outcomes in the in combined data of PREFER in VTE registry and Hokusai-VTE trial..

	n events / n total (%)	Incidence rate (per 100 person-years of follow-up)	Unadjusted HR (95% CI)	Adjusted HR (95% CI) [a]
Recurrent VTE				
Cluster 1	64/3744 (2)	4.7 (1.6-11.7)	0.68 (0.55-1.34)	0.92 (0.65-1.29)
Cluster 2	6/598 (1)	2.4 (0.2-7.2)	0.44 (0.18-1.07)	0.27 (0.12-0.61)
Cluster 3	48/2665 (2)	4.8 (1.6-11.7)	0.87 (0.54-1.40)	1.29 (0.92-1.81)
Cluster 4	27/1367 (2)	5.5 (2.2-13.1)	Ref	Ref
Cluster 5	8/607 (1)	4.0 (1.1-10.2)	0.73 (0.33-1.61)	0.70 (0.37-1.30)
Bleeding				
Cluster 1	31/3744 (1)	2.3 (0.2-7.2)	0.29 (0.18-0.46)	0.50 (0.38-0.66)
Cluster 2	3/598 (1)	1.2 (0.0-5.6)	0.16 (0.05-0.51)	0.23 (0.11-0.46)
Cluster 3	39/2665 (1)	3.9 (1.1-10.2)	0.49 (0.32-0.77)	0.55 (0.41-0.73)
Cluster 4	39/1367 (3)	8.0 (3.5-15.8)	Ref	Ref
Cluster 5	13/608 (2)	6.5 (2.2-13.1)	0.79 (0.42-1.48)	1.21 (0.86-1.69)
Mortality				
Cluster 1	18/3744 (0)	1.4 (0.0-5.6)	0.21 (0.12-0.37)	0.32 (0.22-0.46)
Cluster 2	1/598 (0)	0.4 (0.0-3.7)	0.06 (0.01-0.48)	0.04 (0.01-0.23)
Cluster 3	37/3665 (1)	3.8 (1.1-10.2)	0.59 (0.36-0.95)	0.66 (0.47-0.93)
Cluster 4	30/1367 (2)	6.4 (2.2-13.1)	Ref	Ref
Cluster 5	4/608 (1)	2.1 (0.2-7.2)	0.31 (0.11-0.88)	0.25 (0.10-0.59)

a Adjusted for extended anticoagulant treatment using inverse probability weighing
Abbreviations: CI, confidence interval; HR, hazard ratio

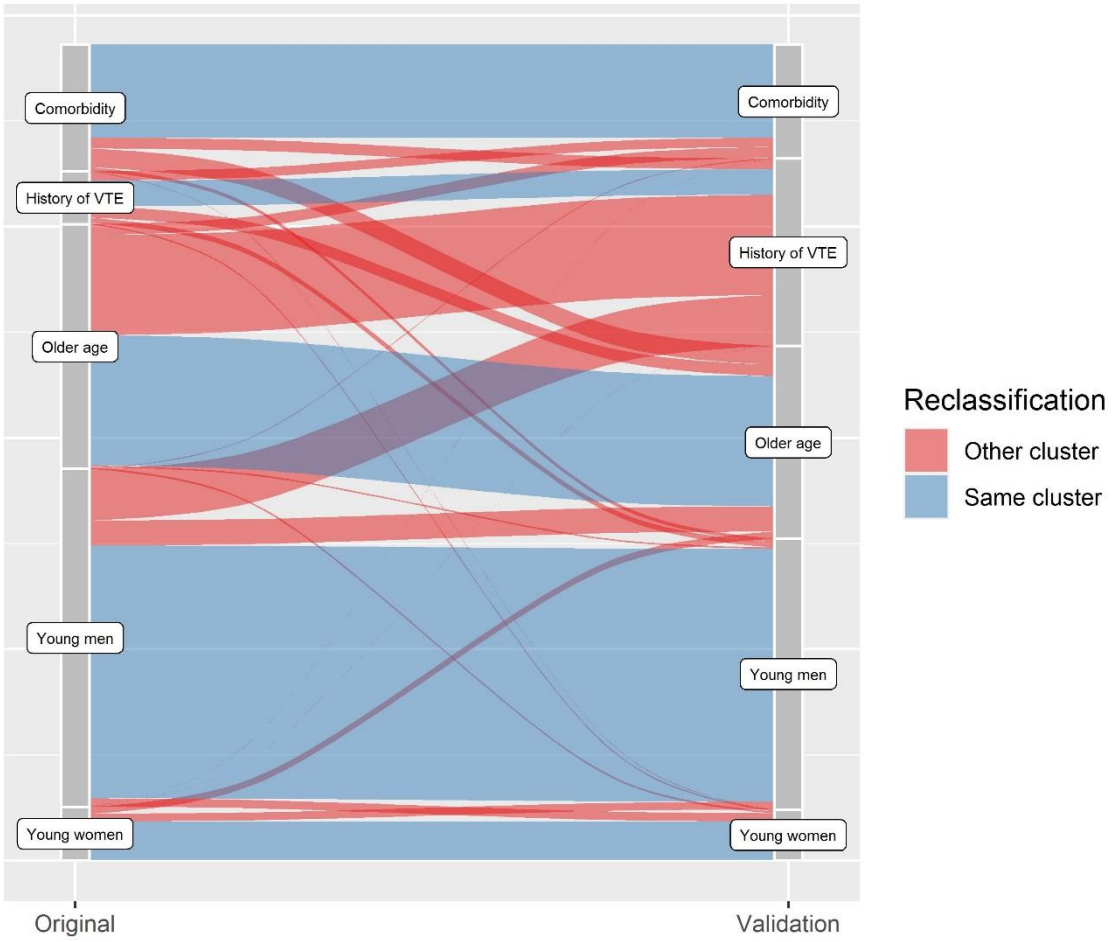
Figures

Figure 1. Novel clusters among patients with venous thromboembolism



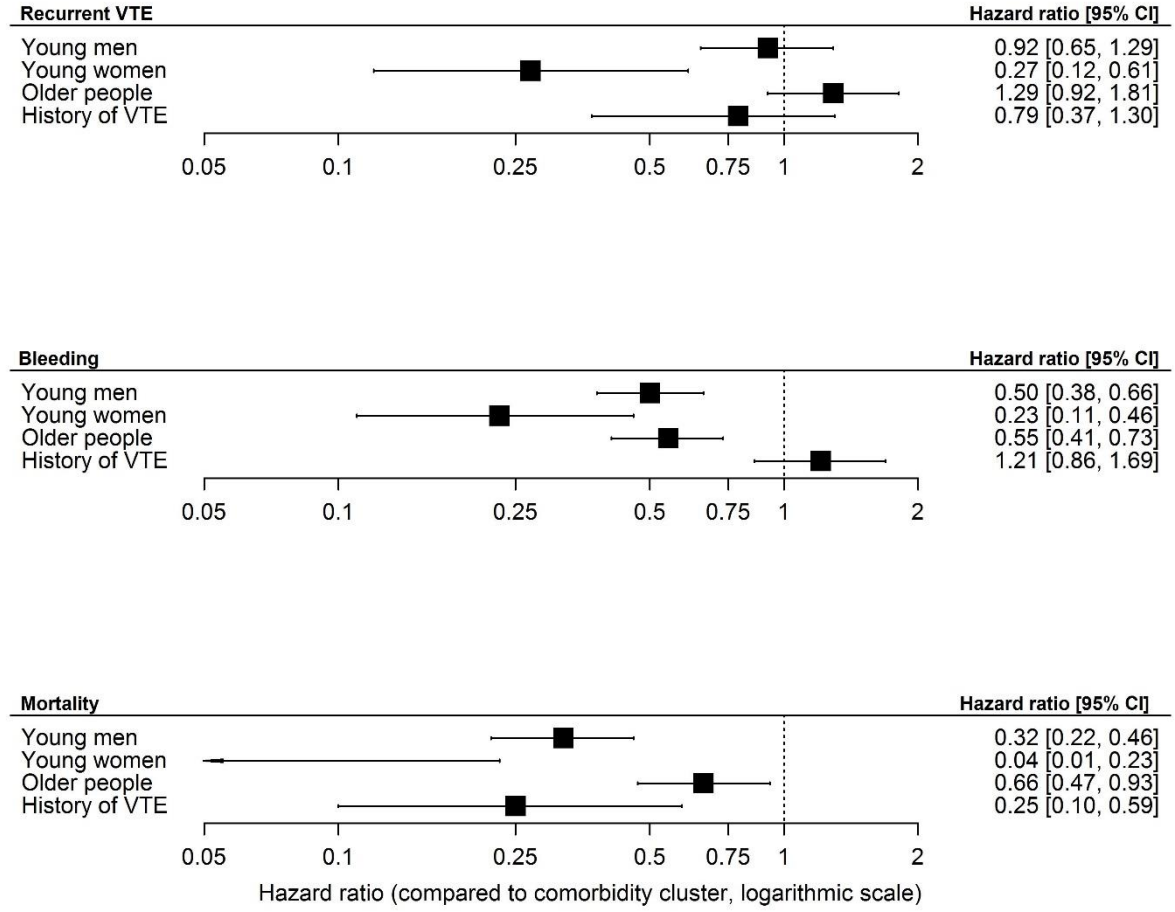
Abbreviations: CVD, cardiovascular disease; DM, diabetes mellitus; PE, pulmonary embolism; VTE, venous thromboembolism

Figure 2. Alluvial plot showing the crossover between clusters derived in the PREFER in VTE registry and clusters derived in Hokusai-VTE in combined data of PREFER in VTE registry and Hokusai-VTE.



Colored splines represent the proportion of patients from each cluster that remain in the same cluster with both the PREFER in VTE registry-cluster model and the Hokusai-VTE cluster model (red) or are redistributed to an alternative cluster (blue). The figure was constructed using combined estimates of the PREFER in VTE registry and the Hokusai-VTE study.

Figure 3. Association between cluster membership and clinical outcomes after adjusting for anticoagulant treatment in combined data of PREFER in VTE registry and Hokusai-VTE trial.



Abbreviations: CI, confidence interval; VTE, venous thromboembolism

Table of contents

References

1 Kearon C, Akl EA, Ornelas J, Blaivas A, Jimenez D, Bounameaux H, Huisman M, King CS, Morris TA, Sood N, Stevens SM, Vintch JRE, Wells P, Woller SC, Moores L. Antithrombotic Therapy for VTE Disease. *Chest* 2016; **149**: 315–52.

2 Konstantinides S V, Meyer G, Becattini C, Bueno H, Geersing G-J, Harjola V-P, Huisman M V, Humbert M, Jennings CS, Jiménez D, Kucher N, Lang IM, Lankeit M, Lorusso R, Mazzolai L, Meneveau N, Ní Áinle F, Prandoni P, Pruszczyk P, Righini M, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). *Eur Heart J* England; 2019; : ehz405.

3 Kearon C, Ageno W, Cannegieter SC, Cosmi B, Geersing G-J, Kyrle PA. Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH. *J Thromb Haemost* England; 2016; **14**: 1480–3.

4 Albertsen IE, Nielsen PB, Søgaard M, Goldhaber SZ, Overvad TF, Rasmussen LH, Larsen TB. Risk of Recurrent Venous Thromboembolism: A Danish Nationwide Cohort Study. *Am J Med* Elsevier Inc.; 2018; **131**: 1067-1074.e4.

5 Ageno W, Farjat A, Haas S, Weitz JI, Goldhaber SZ, Turpie AGG, Goto S, Angchaisuksiri P, Dalsgaard Nielsen J, Kayani G, Schellong S, Bounameaux H, Mantovani LG, Prandoni P, Kakkar AK. Provoked versus unprovoked venous thromboembolism: Findings from GARFIELD-VTE. *Res Pract Thromb Haemost* 2021; **5**: 326–41.

6 Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet (London, England)* England; 1999; **353**: 1167–73.

7 Ahmad T, Pencina MJ, Schulte PJ, O'Brien E, Whellan DJ, Piña IL, Kitzman DW, Lee KL, O'Connor CM, Felker GM. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol* 2014; **64**: 1765–74.

8 Seymour CW, Kennedy JN, Wang S, Chang CCH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H, Huang DT, Kellum JA, Mi Q, Opal SM, Talisa V, Van Der Poll T, Visweswaran S, Vodovotz Y, Weiss JC, Yealy DM, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA - J Am Med Assoc* 2019; **321**: 2003–17.

9 Geiser C, Lehmann W, Eid M. Separating "Rotators" From "Nonrotators" in the Mental Rotations Test: A Multigroup Latent Class Analysis. *Multivariate Behav Res* United States; 2006; **41**: 261–93.

10 Inohara T, Shrader P, Pieper K, Blanco RG, Thomas L, Singer DE, Freeman J V., Allen LA, Fonarow GC, Gersh B, Ezekowitz MD, Kowey PR, Reiffel JA, Naccarelli G V., Chan PS, Steinberg BA, Peterson ED, Piccini JP. Association of of atrial fibrillation clinical phenotypes with treatment patterns and outcomes a multicenter registry study. *JAMA Cardiol* 2018; **3**: 54–63.

11 Uijl A, Savarese G, Vaartjes I, Dahlström U, Brugts JJ, Linssen GC, Empel V, Rocca HB, Asselbergs FW, Lund LH, Hoes AW, Koudstaal S. Identification of Distinct Phenotypic Clusters in Heart Failure with Preserved Ejection Fraction. *Eur J Heart Fail* 2021; **23**: 973–82.

12 Cohen AT, Gitt AK, Bauersachs R, Fronk EM, Laeis P, Mismetti P, Monreal M, Willich SN, Bramlage P, Agnelli G. The management of acute venous thromboembolism in clinical practice results from the european prefer in vte

registry. *Thromb Haemost* 2017; **117**: 1326–37.
13 Büller HR, Décousus H, Grosso MA, Mercuri M, Middeldorp S, Prins MH, Raskob GE, Schellong SM, Schwocho L, Segers A, Shi M, Verhamme P, Wells P. Edoxaban versus warfarin for the treatment of symptomatic venous thromboembolism. *N Engl J Med* 2013; **369**: 1406–15.
14 Linzer DA, Lewis JB. poLCA: An R package for polytomous variable latent class analysis. *J Stat Softw* 2011; **42**: 1–29.
15 Mitchell JD, Gage BF, Fergestrom N, Novak E, Villines TC. Inverse Probability of Treatment Weighting (Propensity Score) using the Military Health System Data Repository and National Death Index. *J Vis Exp* 2020; .
16 Konstantinides S V, Torbicki A, Agnelli G, Danchin N, Fitzmaurice D, Galiè N, Gibbs JSR, Huisman M V, Humbert M, Kucher N, Lang I, Lankeit M, Lekakis J, Maack C, Mayer E, Meneveau N, Perrier A, Pruszczyk P, Rasmussen LH, Schindler TH, et al. 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J* England; 2014; **35**: 3033-3069k.
17 Tritschler T, Castellucci LA, van Es N, Aujesky D, Le Gal G. Treatment of venous thromboembolism in elderly patients in the era of direct oral anticoagulants. *Polish Arch Intern Med* 2020; **130**: 529–38.
18 Roach REJ, Cannegieter SC, Lijfering WM. Differential risks in men and women for first and recurrent venous thrombosis: The role of genes and environment. *J Thromb Haemost* 2014; **12**: 1593–600.
19 Eischer L, Eichinger S, Kyrle PA. The risk of recurrence in women with venous thromboembolism while using estrogens: A prospective cohort study. *J Thromb Haemost* 2014; **12**: 635–40.
20 Skeith L, Le Gal G, Rodger MA. Oral contraceptives and hormone replacement therapy: How strong a risk factor for venous thromboembolism? *Thromb Res* 2021; **202**: 134–8.
21 Kyrle PA, Minar E, Bialonczyk C, Hirschl M, Weltermann A, Eichinger S. The Risk of Recurrent Venous Thromboembolism in Men and Women. *N Engl J Med* Massachusetts Medical Society; 2004; **350**: 2558–63.
22 Douketis J, Tosetto A, Marcucci M, Baglin T, Cosmi B, Cushman M, Kyrle P, Poli D, Tait RC, Iorio A. Risk of recurrence after venous thromboembolism in men and women: Patient level meta-analysis. *Bmj* 2011; **342**: 535.
23 Kiconco S, Abdul Sultan A, Grainge MJ. Recurrence risk of venous thromboembolism and hormone use in women from England: a cohort study using clinical practice research datalink. *Br J Haematol* England; 2017; **177**: 127–35.
24 Ageno W, Becattini C, Brighton T, Selby R, Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism: A meta-analysis. *Circulation* 2008; **117**: 93–102.
25 Schmidt M, Cannegieter SC, Johannesdottir SA, Dekkers OM, Horváth-Puhó E, Sørensen HT. Statin use and venous thromboembolism recurrence: A combined nationwide cohort and nested case-control study. *J Thromb Haemost* 2014; **12**: 1207–15.
26 Ray JG, Rosendaal FR. The role of dyslipidemia and statins in venous thromboembolism. *Curr Control Trials Cardiovasc Med* Department of Medicine, University of Toronto, Toronto, and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. rayjg@mcmaster; 2001; **2**: 165–70.

27 Lange N, Méan M, Stalder O, Limacher A, Tritschler T, Rodondi N, Aujesky D. Anticoagulation quality and clinical outcomes in multimorbid elderly patients with acute venous thromboembolism. *Thromb Res* Elsevier; 2019; **177**: 10–6.
28 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* Elsevier Ltd; 2021; **199**: 85–96.
29 Zuily S, Clerc-Urmès I, Bauman C, Andrade D, Sciascia S, Pengo V, Tektonidou MG, Ugarte A, Gerosa M, Michael Belmont H, Zamorano MAA, Fortin P, Ji L, Efthymiou M, Cohen H, Branch DW, Jesus GR de, Nalli C, Petri M, Rodriguez E, et al. Cluster analysis for the identification of clinical phenotypes among antiphospholipid antibody-positive patients from the APS ACTION Registry. *Lupus* 2020; .
30 Naymagon L, Zubizarreta N, Feld J, Gerwen M Van, Alsen M, Thibaud S, Kessler A, Venugopal S, Makki I, Qin Q, Dharmapuri S, Jun T, Bhalla S, Berwick S, Christian K. Admission D-dimer levels, D-dimer trends, and outcomes in COVID-19 Leonard. *Thromb Res* 2020; **196**: 99–105.
31 Gao Y, Wang N, Li Y, Huang H, Jia D. Phenotyping of non-high-risk acute pulmonary embolism patients: different initial manifestations of clinical deterioration. *Eur J Intern Med* 2021; **90**: 116–8.

Appendix

Expanded methods

Cluster model derivation

Novel clusters with similar clinical characteristics were identified by means of latent class analysis (LCA) using the *poLCA* package (R Statistical Software).[1] LCA assumes the existence of latent classes within a population, explaining patterns of association between clinical characteristics. This method was chosen because it is suitable for analyzing frequently used categorical clinical variables and it has been used in studies with similar research questions in the cardiovascular field.[2,3] As LCA is suitable only for analysis of categorical variables, continuous variables were categorized based on international consensus.[1] Age was analyzed per decade. Maximum likelihood estimation over 10 iterations was used to identify the most common patterns of the clinical variables (as listed above) for a number of clusters. Using an unsupervised approach, no outcomes were included for LCA itself[4]. Each patient had a certain probability of belonging to each cluster. The most likely cluster for each patient was decided based on the sum of partial probabilities of membership per cluster for every variable used in the LCA. The process was repeated with different numbers of clusters (2-10). The optimal number was determined using the first minima of the Bayesian Information Criterion (BIC) and likelihood ratio test. Three models with the lowest BIC were further inspected. Among these, the option resulting in the most meaningful subgroups from a clinical perspective was chosen based on a consensus discussion between authors (JD, MN, AU, MW). To compare characteristics among clusters, Chi-square, ANOVA, and Kruskal-Wallis tests were used.

External validation

The probabilities for each cluster per variable derived in PREFER in VTE were applied to the Hokusai-VTE data, to calculate cluster probabilities for each patient using the original cluster model definition. Patients were then classified according to the highest probability of cluster membership. Median and interquartile range of probabilities were compared among the validation and derivation datasets. To better study the reproducibility of our findings in external data, all steps of cluster model development were repeated in the Hokusai-VTE study. Clusters and probabilities for cluster membership between the original cluster model and the cluster model derived in Hokusai-VTE were assessed and compared. Subsequently, the proportion of patients being classified as belonging to the same cluster according to both the original cluster model and the cluster model developed in Hokusai-VTE was assessed to demonstrate the reproducibility of our findings. These results are shown graphically using an alluvial plot.

Association between cluster membership and clinical outcomes

To assess the association between cluster membership and clinical outcomes after initial anticoagulant treatment, Cox proportional hazard models were fitted with cluster membership as a categorical covariate, using time to recurrent venous thromboembolism (VTE), clinically relevant bleeding, and all-cause mortality in separate analyses as outcome variables. Combined data from the Hokusai-VTE trial and PREFER in VTE registry was used for the main analysis, to ensure sufficient numbers of outcomes per cluster. Analysis were repeated in the Hokusai-VTE trial and PREFER in VTE registry separately. . Inverse probability of treatment weighting (IPTW) was used to account for possible confounding by indication. With IPTW, patients are weighted by their probability of receiving treatment using a propensity score.[5] This approach is flexible and does not reduce precision as in the case of propensity score matching.[6] Propensity scores were calculated using a logistic regression analysis with treatment status as dependent variable and factors associated with both treatment assignment and the outcome of interest as predictors. Separate propensity scores were calculated for recurrent VTE, clinically relevant bleeding, and mortality, since variables should be associated with the outcome of interest as well as with the exposure, to reduce bias and decrease variance.[6,7] Variables for the propensity scores were selected based on available literature, including prediction models for recurrent VTE and bleeding, and availability in the Hokusai-VTE trial (Table 1).[8–12] Analyses were repeated after truncating propensity scores at the 1st and 99th percentiles to reduce the possibility of overrepresenting patients in atypical circumstances with risks that were too high. This did not influence the results. The proportional hazards assumption for these Cox models was tested by visually inspecting Schoenfeld residuals. Hazard ratios (HRs) and 95% confidence intervals (95% CIs) were reported for clusters 1, 2, 3, and 5 compared to the reference cluster 4, which was selected as the cluster with the highest event rates of recurrent VTE and bleeding.

Appendix Table S1. Definitions of variables used in the latent class analysis and definitions of outcomes.

	PREFER in VTE	**Hokusai-VTE**
Age	In decades	In years
Sex	Male, female	Male, female
Site of index event	PE, DVT or both	PE, DVT or both
Provoking factors		
Estrogen therapy	Hormone replacement therapy or oral contraceptives at the time of the index event	Hormone replacement therapy or oral contraceptives at the time of the index event
Surgery, trauma, or immobilization	Major surgery, trauma, prolonged immobilization or >5 days in bed <90 days preceding the index VTE event	Any surgery, trauma or immobilization <90 days preceding the index VTE event
Medical history		
Prior stroke	Ischemic or hemorrhagic stroke or TIA	Ischemic or hemorrhagic stroke or TIA
History of bleeding	History of major or clinically relevant non-major bleeding according to the ISTH definitions [13,14]	History of life-threatening bleeding or gastrointestinal bleeding
Hypertension	SBP >130/80 mmHg in patients with diabetes; SBP >140 mmHg or DBP >90 mmHg in those without diabetes	History of hypertension
Cardiovascular disease	Documented coronary artery disease or vascular disease	Stable angina pectoris, coronary artery disease, coronary or carotid artery occlusion or stenosis were interpreted as "unequivocally documented ASCVD; occlusion or stenosis of not further specified artery and 'insufficiency' or 'disease' were not considered unequivocally documented. Renal artery stenosis, aortic thrombosis and aortic dissection, were not considered cardiovascular disease (as these may have other causes)
Congestive heart failure	Congestive heart failure or left ventricular systolic dysfunction	Symptomatic heart failure, present or in medical history
Known thrombophilia	Antithrombin deficiency, protein C, or protein S, factor V Leiden or prothrombin gene mutation, and/or antiphospholipid syndrome	Antithrombin deficiency, protein C, or protein S, factor V Leiden or prothrombin gene mutation, and/or antiphospholipid syndrome
Previous VTE	Objectively confirmed PE or DVT before the index event	Previous PE or DVT
Chronic lung disease	Any chronic respiratory disease	Any chronic lung disease (e.g. a1-AT deficiency, asthma, atelectasis/bronchiectasis, COPD, emphysema, ILD, fibrosis, pulmonary hypertension)
Diabetes mellitus	Type I or type II	Type I or type II
Renal disease	Chronic kidney disease (GFR<15 and/or end-stage renal disease)	Chronic kidney disease (GFR<15 and/or end-stage renal disease)
History of cancer	Any previous cancer diagnosis not considered active and without ongoing treatment	Cancer not diagnosed or receiving treatment within 6 months prior to the index event or metastatic cancer; excluding non-melanoma skin cancer
Active cancer [a]	Active cancer with ongoing treatment was noted on case report forms	Any cancer excluding non-melanoma skin cancer diagnosed or receiving cancer treatment within 6 months prior to the index VTE event, or metastatic cancer
Medication use		
Antiplatelet therapy	Chronic treatment with aspirin or platelet inhibitors, indicated at the time of the index event	Concomitant use of aspirin, platelet inhibitors or other antithrombotic agents during study follow-up
Laboratory measurements		
Thrombocytopenia	Platelet count $<50*10^9$/L	Platelet count $<50*10^9$/L
Outcomes		

Recurrent VTE	Objectively confirmed recurrent DVT or PE	Recurrent symptomatic, adjudicated DVT or PE including death in which PE could not be ruled out as potential cause
Clinically relevant bleeding	Composite of major and clinically relevant, non-major bleeding according to the ISTH definitions [13,14]	Composite of major and clinically relevant, non-major bleeding according to the ISTH definitions [13,14]
Death	All-cause mortality	All-cause mortality

Abbreviations: ASCVD atherosclerotic cardiovascular disease; DBP, diastolic blood pressure; DVT, deep venous thrombosis; GFR, glomerular filtration rate; ISTH, International Society on Thrombosis and Haemostasis; PE, pulmonary embolism; SBP, systolic blood pressure; TIA, transient ischaemic attack; VTE, venous thromboembolism

a Active cancer was not used for latent class analysis, but as an exclusion criterion for the present study

Appendix Table S2. Probabilities for each classification per cluster.

All data are shown as median (interquartile range) probability of patients in each cluster to be in any of the other groups. Probabilities are based on the original cluster model, developed in the PREFER in VTE registry.

	Probability for cluster 1 (young men)	**Probability for cluster 2** (young women)	**Probability for cluster 3** (older age)	**Probability for cluster 4** (comorbidity)	**Probability for cluster 5** (history of VTE)
PREFER in VTE					
Cluster 1	**0.86 (0.71-0.97)**	0.00 (0.00-0.04)	0.00 (0.00-0.02)	0.00 (0.00-0.00)	0.04 (0.01-0.17)
Cluster 2	0.00 (0.00-0.00)	**1.00 (1.00-1.00)**	0.00 (0.00-0.00)	0.00 (0.00-0.00)	0.00 (0.00-0.00)
Cluster 3	0.01 (0.00-0.07)	0.00 (0.00-0.00)	**0.70 (0.53-0.88)**	0.03 (0.01-0.11)	0.03 (0.00-0.33)
Cluster 4	0.00 (0.00-0.00)	0.00 (0.00-0.00)	0.00 (0.00-0.06)	**0.95 (0.77-1.00)**	0.01 (0.00-0.11)
Cluster 5	0.02 (0.00-0.14)	0.00 (0.00-0.00)	0.02 (0.01-0.05)	0.03 (0.02-0.06)	**0.86 (0.72-0.91)**
Hokusai-VTE					
Cluster 1	**0.87 (0.70-0.97)**	0.00 (0.00-0.00)	0.04 (0.01-0.16)	0.00 (0.00-0.00)	0.00 (0.00-0.02)
Cluster 2	0.00 (0.00-0.00)	**1.00 (1.00-1.00)**	0.00 (0.00-0.00)	0.00 (0.00-0.00)	0.00 (0.00-0.00)
Cluster 3	0.05 (0.00-0.20)	0.00 (0.00-0.00)	**0.82 (0.65-0.90)**	0.02 (0.01-0.05)	0.03 (0.01-0.06)
Cluster 4	0.00 (0.00-0.00)	0.00 (0.00-0.00)	0.06 (0.00-0.15)	**0.83 (0.60-0.98)**	0.05 (0.00-0.19)
Cluster 5	0.02 (0.00-0.11)	0.00 (0.00-0.00)	0.04 (0.00-0.18)	0.07 (0.02-0.21)	**0.69 (0.53-0.84)**

Abbreviations: VTE venous thromboembolism

Appendix Table S3. Probabilities for each variable per cluster.

Probabilities are based on the original cluster model, developed in the PREFER in VTE registry.

	Cluster 1 (young men)	**Cluster 2** (young women)	**Cluster 3** (older age)	**Cluster 4** (comorbidity)	**Cluster 5** (history of VTE)
Age (years)					
<20	0.010	0.027	0.000	0.000	0.000
20-29	0.079	0.272	0.000	0.000	0.000
30-39	0.159	0.271	0.006	0.005	0.000
40-49	0.273	0.359	0.029	0.016	0.088
50-59	0.259	0.051	0.140	0.077	0.096
60-69	0.169	0.016	0.282	0.199	0.389
70-79	0.048	0.004	0.343	0.319	0.341
80-89	0.000	0.000	0.180	0.353	0.072
>90	0.000	0.000	0.021	0.032	0.013
Female sex	0.317	0.991	0.514	0.498	0.391
Index event: PE	0.326	0.382	0.468	0.487	0.338
Provoking factors					
Estrogen therapy	0.000	0.764	0.009	0.002	0.007
Surgery, trauma, or immobilization	0.358	0.347	0.308	0.472	0.074
Medical history					
Prior stroke	0.000	0.011	0.013	0.252	0.000
History of bleeding	0.026	0.020	0.031	0.089	0.000
Hypertension	0.078	0.059	0.640	0.784	0.665
Cardiovascular disease	0.011	0.000	0.016	0.529	0.144
Congestive heart failure	0.000	0.000	0.023	0.203	0.000
Known thrombophilia	0.077	0.083	0.000	0.030	0.256
Previous VTE	0.276	0.069	0.145	0.250	0.745
Chronic lung disease	0.036	0.020	0.102	0.219	0.056
Diabetes mellitus	0.007	0.000	0.138	0.247	0.160
Renal disease	0.000	0.004	0.070	0.167	0.089
History of cancer	0.013	0.030	0.135	0.113	0.137
Medication use					
Antiplatelet therapy	0.003	0.000	0.009	0.573	0.275
Laboratory measurements					
Thrombocytopenia	0.003	0.000	0.007	0.005	0.000

Abbreviations: PE pulmonary embolism; VTE venous thromboembolism

Appendix Table S4. Probabilities for each classification per cluster, based on the cluster model developed in the Hokusai-VTE study.

All data are shown as median (interquartile range) probability of patients in each cluster to be in any of the other groups. Probabilities are based on the cluster model developed in the Hoksuai-VTE study, developed as a method of external validation.

	Probability for cluster 1 (young men)	**Probability for cluster 2** (young women)	**Probability for cluster 3** (older age)	**Probability for cluster 4** (comorbidity)	**Probability for cluster 5** (history of VTE)
PREFER in VTE					
Cluster 1	**0.86 (0.71-0.97)**	0.00 (0.00-0.00)	0.02 (0.00-0.15)	0.00 (0.00-0.00)	0.00 (0.00-0.14)
Cluster 2	0.00 (0.00-0.00)	**0.99 (0.95-1.00)**	0.01 (0.00-0.05)	0.00 (0.00-0.00)	0.00 (0.00-0.00)
Cluster 3	0.01 (0.00-0.01)	0.00 (0.00-0.00)	**0.84 (0.68-0.93)**	0.06 (0.01-0.15)	0.00 (0.00-0.14)
Cluster 4	0.00 (0.00-0.00)	0.00 (0.00-0.00)	0.03 (0.00-0.19)	**0.88 (0.70-0.98)**	0.05 (0.02-0.14)
Cluster 5	0.03 (0.01-0.18)	0.00 (0.00-0.00)	0.02 (0.01-0.04)	0.05 (0.02-0.14)	**0.80 (0.65-0.88)**
Hokusai-VTE					
Cluster 1	**0.86 (0.71-0.97)**	0.00 (0.00-0.00)	0.00 (0.00-0.02)	0.00 (0.00-0.00)	0.03 (0.00-0.18)
Cluster 2	0.00 (0.00-0.00)	**1.00 (0.99-1.00)**	0.00 (0.00-0.00)	0.00 (0.00-0.00)	0.00 (0.00-0.00)
Cluster 3	0.02 (0.00-0.13)	0.00 (0.00-0.00)	**0.87 (0.68-0.93)**	0.05 (0.02-0.1)	0.00 (0.00-0.00)
Cluster 4	0.00 (0.00-0.00)	0.00 (0.00-0.00)	0.01 (0.00-0.06)	**0.88 (0.70-0.98)**	0.02 (0.00-0.17)
Cluster 5	0.08 (0.01-0.22)	0.00 (0.00-0.00)	0.02 (0.01-0.03)	0.03 (0.01-0.10)	**0.78 (0.61-0.88)**

Abbreviations: VTE venous thromboembolism

Appendix Table S5. Probabilities for each variable per cluster for the cluster model developed in the Hokusai-VTE study.

Probabilities are based on the cluster model developed in the Hoksuai-VTE study, developed as a method of external validation.

	Cluster 1 (young men)	**Cluster 2** (young women)	**Cluster 3** (older age)	**Cluster 4** (comorbidity)	**Cluster 5** (history of VTE)
Age (years)					
<20	0.000	0.014	0.000	0.000	0.000
20-29	0.116	0.286	0.004	0.000	0.000
30-39	0.231	0.302	0.013	0.010	0.014
40-49	0.309	0.276	0.093	0.026	0.119
50-59	0.233	0.117	0.161	0.114	0.281
60-69	0.094	0.003	0.305	0.303	0.361
70-79	0.017	0.002	0.313	0.350	0.191
80-89	0.000	0.000	0.108	0.180	0.033
>90	0.000	0.000	0.003	0.016	0.001
Female sex	0.312	1.000	0.958	0.384	0.000
Index event: PE	0.313	0.482	0.425	0.482	0.309
Provoking factors					
Estrogen therapy	0.000	1.000	0.040	0.006	0.000
Surgery, trauma, or immobilization	0.227	0.174	0.248	0.158	0.114
Medical history					
Prior stroke	0.001	0.002	0.015	0.160	0.009
History of bleeding	0.007	0.000	0.008	0.025	0.000
Hypertension	0.047	0.085	0.626	0.828	0.514
Cardiovascular disease	0.002	0.002	0.041	0.408	0.064
Congestive heart failure	0.001	0.000	0.018	0.136	0.011
Known thrombophilia	0.081	0.041	0.031	0.034	0.058
Previous VTE	0.178	0.033	0.194	0.246	0.208
Chronic lung disease	0.049	0.076	0.142	0.245	0.112
Diabetes mellitus	0.019	0.015	0.134	0.261	0.121
Renal disease	0.001	0.000	0.007	0.071	0.010
History of cancer	0.011	0.011	0.119	0.124	0.066
Medication use					
Antiplatelet therapy	0.120	0.094	0.128	0.671	0.129
Laboratory measurements					
Thrombocytopenia	0.000	0.000	0.001	0.001	0.000

Abbreviations: PE pulmonary embolism; VTE venous thromboembolism

Appendix Table S6. Association between cluster membership and clinical outcomes in Hokusai-VTE.

	n events / n total (%)	Incidence rate (per 100 person- years of follow-up)	Unadjusted HR (95% CI)	Adjusted HR (95% CI) [a]
Recurrent VTE				
Cluster 1	40/2866 (1)	4.7 (1.6-11.7)	0.72 (0.42-1.24)	0.75 (0.45-1.26)
Cluster 2	2/420 (0)	1.4 (0.0-5.6)	0.21 (0.05-0.88)	0.20 (0.05-0.79)
Cluster 3	20/1791 (2)	5.4 (1.6-11.7)	0.82 (0.46-1.46)	0.79 (0.45-1.38)
Cluster 4	27/1058 (2)	6.7 (2.8-14.4)	Ref	Ref
Cluster 5	6/458 (1)	5.9 (2.2-13.1)	0.88 (0.35-2.20)	0.91 (0.37-2.20)
Bleeding				
Cluster 1	24/2866 (1)	2.8 (0.6-8.8)	0.24 (0.14-0.40)	0.39 (0.32-0.48)
Cluster 2	3/420 (1)	2.0 (0.2-7.2)	0.18 (0.06-0.58)	0.30 (0.17-0.52)
Cluster 3	24/1791 (1)	4.8 (1.6.11.7)	0.40 (0.24-0.67)	0.21 (0.15-0.29)
Cluster 4	36/1058 (3)	12.3 (6.2-21.0)	Ref	Ref
Cluster 5	10/458 (2)	10.1 (4.8-18.4)	0.75 (0.37-1.51)	1.21 (0.92-1.58)
Mortality				
Cluster 1	13/2866 (0)	1.6 (0.2-7.2)	0.20 (0.10-0.40)	0.21 (0.11-0.41)
Cluster 2	1/420 (0)	0.7 (0.0-5.6)	0.09 (0.01-0.68)	0.08 (0.01-0.60)
Cluster 3	22/1791 (1)	4.5 (1.1-10.2)	0.58 (0.32-1.05)	0.56 (0.31-0.99)
Cluster 4	22/1058 (2)	7.8 (3.5-15.8)	Ref	Ref
Cluster 5	3/458 (1)	3.2 (0.6-8.8)	0.38 (0.11-1.29)	0.40 (0.13-1.28)

a Adjusted for extended anticoagulant treatment using inverse probability weighing
Abbreviations: CI, confidence interval; HR, hazard ratio

Appendix Table S7. Association between cluster membership and clinical outcomes in the PREFER in VTE registry.

	n events / n total (%)	Incidence rate (per 100 person- years of follow-up)	Unadjusted HR (95% CI)	Adjusted HR (95% CI) [a]
Recurrent VTE				
Cluster 1	42/912 (5)	8.2 (3.5-15.8)	0.98 (0.55-1.75)	0.86 (0.60-1.24)
Cluster 2	4/189 (2)	3.9 (1.1-10.2)	0.46 (0.15-1.38)	0.44 (0.22-0.87)
Cluster 3	38/923 (4)	7.4 (2.8-14.4)	0.89 (0.50-1.60)	0.85 (0.59-1.22)
Cluster 4	16/341 (5)	8.2 (3.5-15.8)	Ref	Ref
Cluster 5	3/151 (2)	3.0 (0.6-8.8)	0.39 (0.11-1.34)	0.24 (0.09-0.62)
Bleeding				
Cluster 1	24/912 (3)	4.7 (1.6-11.7)	0.47 (0.26-0.85)	0.38 (0.25-0.56)
Cluster 2	6/189 (3)	6.1 (2.2-13.1)	0.58 (0.23-1.44)	0.47 (0.25-0.88)
Cluster 3	36/923 (4)	6.9 (2.8-14.4)	0.69 (0.40-1.21)	0.58 (0.40-0.83)
Cluster 4	19/341 (6)	9.9 (4.8-18.4)	Ref	Ref
Cluster 5	3/151 (2)	3.0 (0.6-8.8)	0.33 (0.10-1.12)	0.28 (0.12-0.67)
Mortality				
Cluster 1	5/912 (1)	1.0 (0.0-5.6)	0.22 (0.07-0.69)	0.21 (0.11-0.39)
Cluster 2	0/189 (0)	0.0 (0.0-3.7)	n.a.	n.a.
Cluster 3	18/923 (2)	3.6 (1.1-10.2)	0.84 (0.36-1.93)	0.62 (0.38-1.01)
Cluster 4	8/341 (2)	4.2 (1.1-10.2)	Ref	Ref
Cluster 5	1/151 (1)	1.0 (0.0-5.6)	0.22 (0.03-1.76)	0.47 (0.19-1.12)

Appendix Figure S1. Flow diagram of patients included in the present study.

Figure A. PREFER in VTE registry

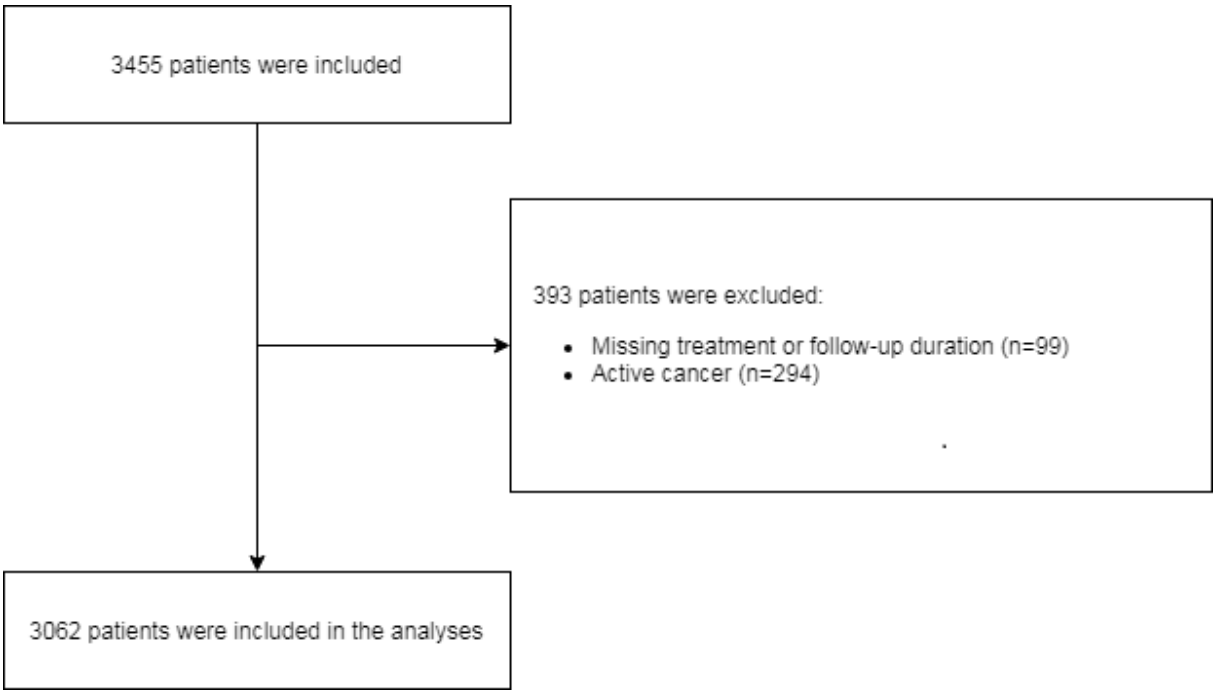
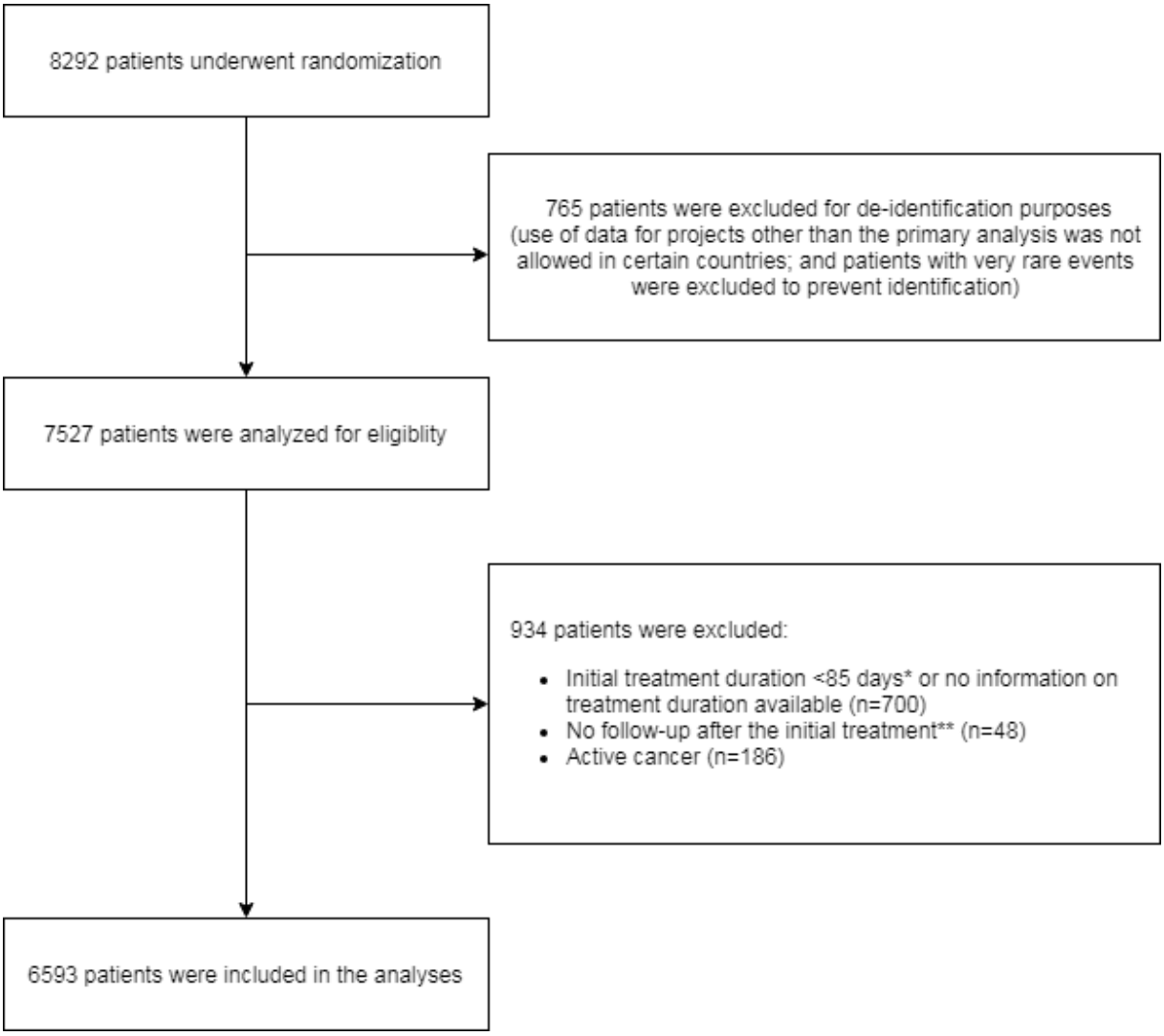
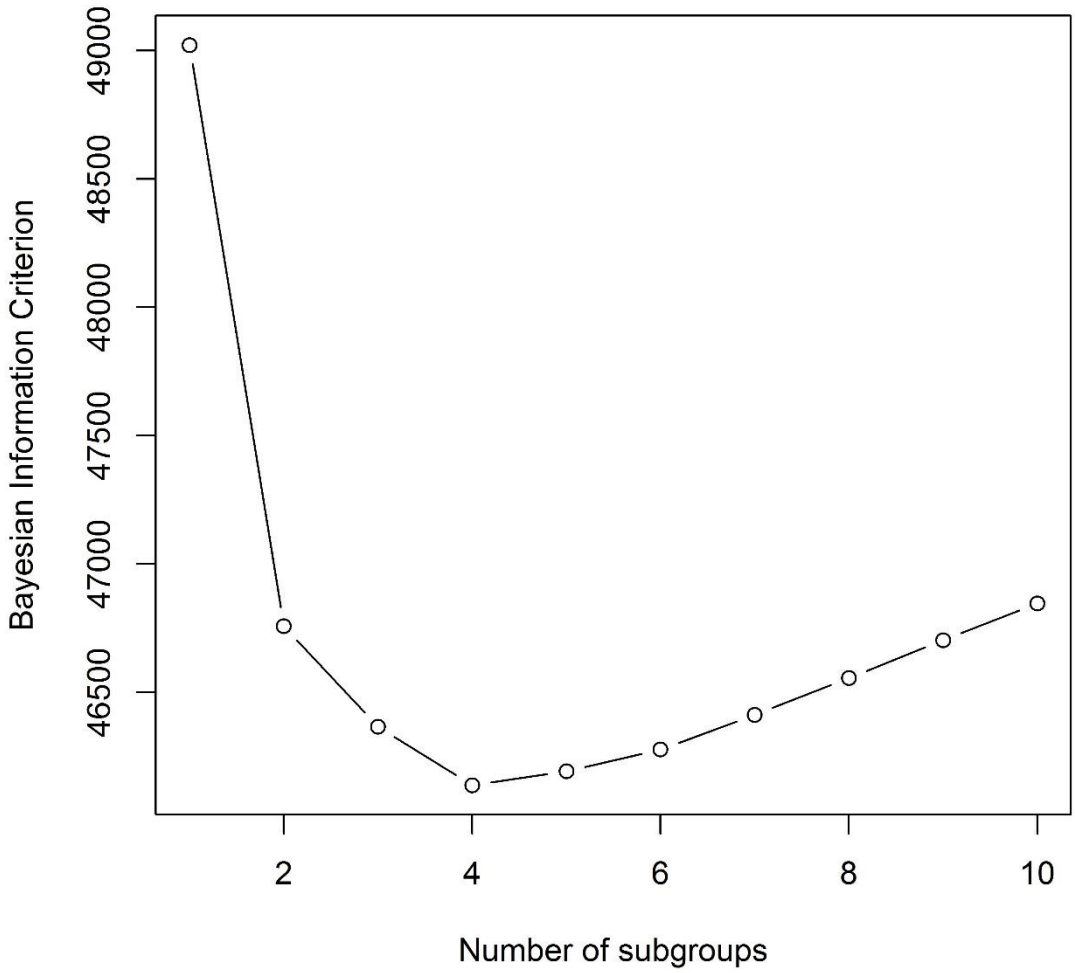


Figure B. Hokusai-VTE

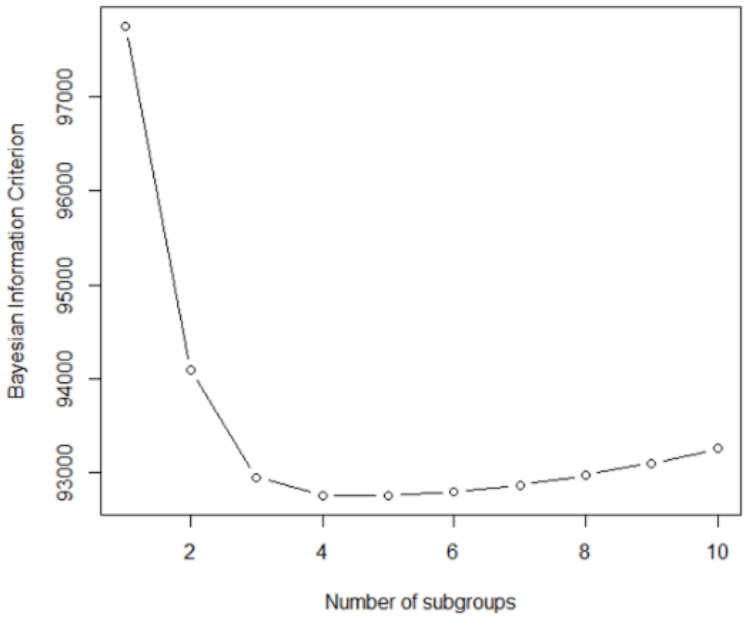


* Corresponding with current international guidelines, initial treatment was assumed to consist of at least 3 months of treatment. As per protocol, follow-up visits were scheduled after 90±5 days. Therefore, 85 days was used as a cut-off point for an initial treatment of at least 3 months.
** Hokusai-VTE enrolled patients at the time of the index event. For the present analyses, follow-up started after an initial treatment of 3-12 months. If treatment was stopped within 12 months, end of treatment was considered start of follow-up for the present analyses. For some patients, this coincided with end of study follow-up. Patients with a total treatment duration >12 months were considered to receive extended treatment.

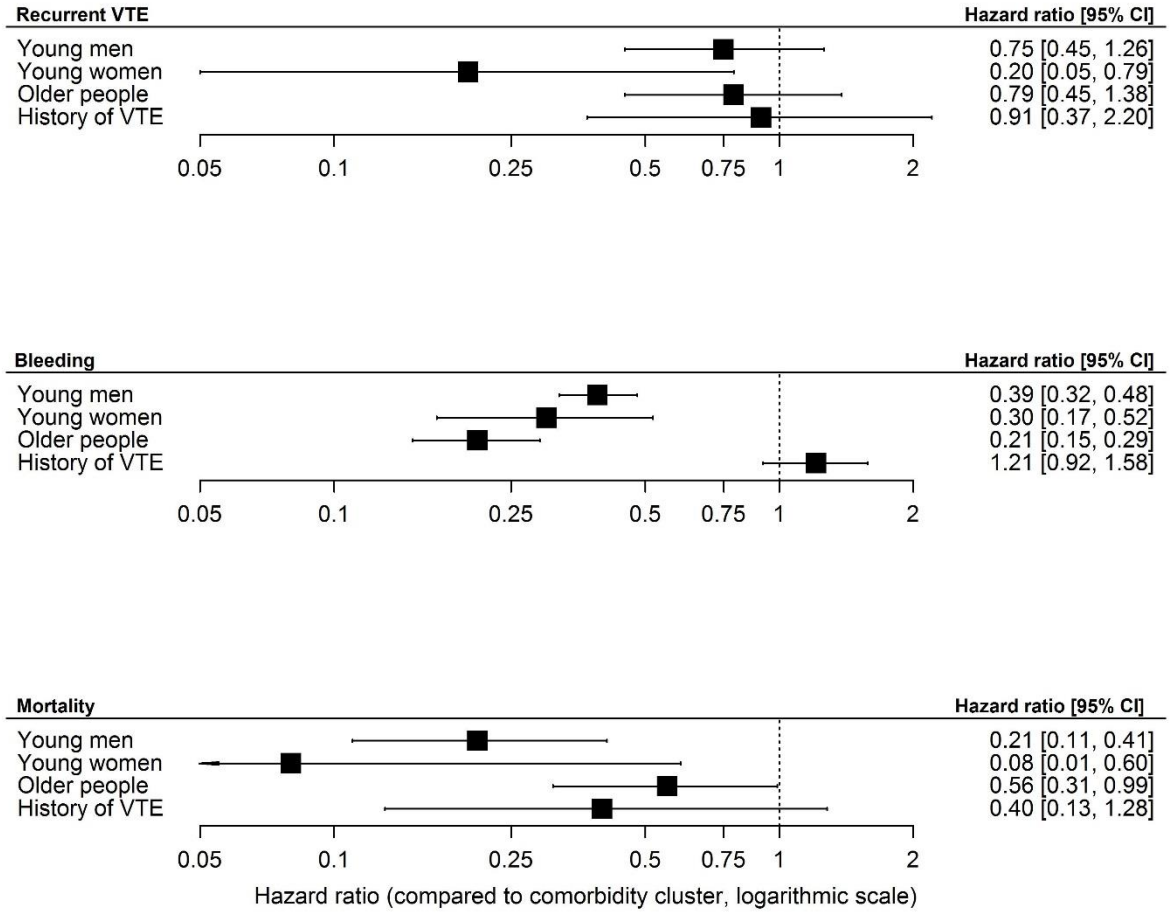
Appendix Figure S2. Comparison of models with 2 to 10 clusters based on the Bayesian Information Criterion in PREFER in VTE.



Appendix Figure S3. Comparison of models with 2 to 10 clusters based on the Bayesian Information Criterion in Hokusai-VTE.



Appendix Figure S4. Association between cluster membership and clinical outcomes after adjusting for anticoagulant treatment in Hokusai-VTE.



Appendix Figure S5. Association between cluster membership and clinical outcomes after adjusting for anticoagulant treatment in PREFER in VTE registry.