

Principal Component Analysis on Recurrent Venous Thromboembolism

Clinical and Applied
Thrombosis/Hemostasis
Volume 25: 1-9
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1076029619895323
journals.sagepub.com/home/cat



Tiago D. Martins, PhD^{1,2}, Joyce M. Annichino-Bizzacchi, PhD³,
Anna V. C. Romano, MSc³, and Rubens Maciel Filho, PhD¹

Abstract

The rates of recurrent venous thromboembolism (RVTE) vary widely, and its causes still need to be elucidated. Statistical multivariate methods can be used to determine disease predictors and improve current methods for risk calculation. The objective of this study was to apply principal component analysis to a set of data containing clinical records of patients with previous venous thromboembolism and extract the main factors that predict recurrent thrombosis. Records of 39 factors including blood and lipid parameters, hereditary thrombophilia, antiphospholipid syndrome, clinical data regarding previous thrombosis and treatment, and Doppler ultrasound results were collected from 235 patients. The results showed that 13 principal components were associated with RVTE and that 18 of 39 factors are the important for the analysis. These factors include red blood cell, white blood cell, hematocrit, red cell distribution width, glucose, lipids, natural anticoagulant, creatinine, age, as well as first deep vein thrombosis data (distal/proximal, D-dimer, and time of anticoagulation). The results demonstrated that simple clinical parameters easy to be collected can be used to predict rates of recurrence and to develop new clinical decision support systems to predict the rates of RVTE.

Keywords

multivariate analysis, statistics, embolism and thrombosis, recurrence, principal component analysis

Date received: 26 September 2019; revised: 01 November 2019; accepted: 18 November 2019.

Introduction

Recurrent venous thromboembolism (RVTE) is an important subject in any medical center. Once the patient is diagnosed with a first venous thromboembolism (VTE), anticoagulation therapy is the prevention method of recurrence commonly used. However, the duration of anticoagulation therapy is a difficult decision and depends on multiple factors that interact with each other and need to be evaluated individually.¹⁻⁵ Dash,⁶ Vienna,⁷ and Men and HERDOO-2⁸ are 3 scores used to decide about anticoagulant prophylaxis by the risk of recurrence, and each one takes different factors in consideration.

Recent reviews showed that these score models have a strong limitation when predicting patients with low risk of RVTE and still lacks a complete validation.^{9,10} Patients with provoked VTE are also not included in these models. There are some studies indicating that some patients with provoked VTE could benefit of prolonged prophylaxis¹¹⁻¹³ and also that they have the same risk of recurrence when comparing those with a previous unprovoked event.¹⁴ The scores also does not include patients with antiphospholipid antibody syndrome, cancer, natural anticoagulant activity deficiencies, and so on.^{9,10}

Thus, it is important for the development of prediction scores that include those individuals. To overcome these issues, sophisticated multivariate statistical techniques could be used to identify the main risk factors of RVTE for a wide range of patients.

Principal component analysis (PCA) is a multivariate statistical method that identifies patterns and classifies the factors that influences a given phenomenon. It is a technique widely

¹ School of Chemical Engineering, University of Campinas, Campinas, Brazil

² Departamento de Engenharia Química, Universidade Federal de São Paulo, Instituto de Ciências Ambientais, Químicas e Farmacêuticas, Diadema, São Paulo, Brazil

³ Hematology and Hemotherapy Center, University of Campinas/ Hemocentro-Unicamp, Instituto Nacional de Ciência e Tecnologia do Sangue, Campinas, São Paulo, Brazil

Corresponding Author:

Tiago D. Martins, Departamento de Engenharia Química, Universidade Federal de São Paulo, Instituto de Ciências Ambientais, Químicas e Farmacêuticas, Rua São Nicolau, 210, Centro, Diadema, São Paulo, CEP 09920-000, Brazil.
Email: tdmartins@unifesp.br



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use,

reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Table 1. Factors Considered in the Present Study.

Sex (men/women)	G20210A prothrombin gene mutation (No/Yes)	Red blood cell (RBC) count, μL^{-1}	Fasting blood glucose, $\text{mg}\cdot\text{dL}^{-1}$
Age, years	Factor VIII (IU· dL^{-1})	Hemoglobin (HB) ($\text{g}\cdot\text{dL}^{-1}$)	Creatinine ($\text{mg}\cdot\text{dL}^{-1}$)
Pulmonary embolism (No/Yes)	Protein C level (PC) ($\text{mg}\cdot\text{dL}^{-1}$)	Hematocrit (HCT) (%)	C-reactive protein ($\text{mg}\cdot\text{dL}^{-1}$)
Side of inferior member (left/right)	Protein S level (PS) ($\text{mg}\cdot\text{dL}^{-1}$)	Red blood cell distribution (RDW) (%)	Arterial hypertension (No/Yes)
Location of VTE in leg (distal/proximal)	Antithrombin level (AT) ($\text{mg}\cdot\text{dL}^{-1}$)	Platelet (PLT) (μL^{-1})	Diabetes mellitus (No/Yes)
Provoked or unprovoked	Antiphospholipid syndrome (No/Yes)	Mean platelet volume (MVP) (fL)	Dyslipidemia (No/Yes)
Anticoagulation time, months	Body mass index (BMI)	Total cholesterol ($\text{mg}\cdot\text{dL}^{-1}$)	Renal failure (No/Yes)
Anticoagulant use (No/Yes)	Smoking (No/Yes)	High-density lipoprotein (HDL) ($\text{mg}\cdot\text{dL}^{-1}$)	Cancer (No/Yes)
D-dimer, $\text{ng}\cdot\text{mL}^{-1}$	Hormonal therapy (for women) (No/Yes)	Low-density lipoprotein (LDL) ($\text{mg}\cdot\text{dL}^{-1}$)	Residual venous thrombus on US Doppler (No/Yes)
FV Leiden (No/Yes)	White blood cell (WBC) count (μL^{-1})	Triglycerides ($\text{mg}\cdot\text{dL}^{-1}$)	

Abbreviations: FV, factor V; US, ultrasound; AT, Antithrombin; PS, Protein S.

used to identify patterns in the medical field.¹⁵⁻¹⁹ Principal component analysis is a decomposition of a set of correlated variables into a set of uncorrelated variables, named as principal components (PCs), which are organized by descending order of variance. Principal component analysis can also be used to reduce dimensionality by cutting the PCs that are less important (less variance) than the original data. The remaining PCs are useful to develop new models, and their loadings can be used to calculate the contribution of each factor in the collected data.

The aim of this study was to collect clinical data of several patients diagnosed with a first thrombotic episode and use PCA to identify the predictor factors for RVTE. To obtain a more comprehensive view of the risk factors, we included patients with provoked and unprovoked VTE, antiphospholipid antibody syndrome, cancer, and natural anticoagulant activity deficiencies.

Methods

Population

In this study, 261 consecutive patients with a first acquired or provoked VTE, assisted at outpatient clinic of Hemocentro Unicamp between January 2009 and August 2016, were followed. Mathematically, the occurrence of RVTE (or not) can be considered as a function of several risk factors. Acquired risk factors for VTE were pregnancy and postpartum, hormone therapy for contraception or hormonal replacement, surgery, trauma, air travel, hereditary thrombophilia, cancer, and antiphospholipid antibodies. Exclusion criteria were patients with missing data, thrombosis that occurred in other sites than pulmonary embolism (PE), lower limbs or central nervous system, and age under 18 years. After the first VTE, all patients were treated with anticoagulant therapy during the standard 6-month

period. Then, every 6 months, each patient returns to the clinical center for check-up examinations.

Recurrent VTE was defined by the detection of a novel thrombotic episode at a different site from the primary event or a new episode occurred at the same localization of previous thrombosis. Thrombosis of the limbs was confirmed by duplex ultrasonography, thrombosis of the cerebral veins by magnetic resonance imaging, and PE by ventilation-perfusion scanning or magnetic resonance imaging.

Table 1 shows all the evaluated factors, which included clinical, thrombotic data, acquired and inherited risk factors for thrombosis, laboratorial parameters, and residual vein thrombosis detected by ultrasound Doppler. All patient data, blood samples, and examinations were collected between the first and the second thrombotic events in an average time of 1 month after the end of anticoagulation therapy. The 39 factors presented in Table 1 were used to perform the PCA. Since there are some factors that are not numerical data, they were converted to a numerical form as shown in Table 2. All numerical factors were used as is.

Principal Component Analysis

The PCs are a new set of variables that are organized in order of its capacity to describe a given phenomenon. To calculate the PCs from an original data matrix \mathbf{X} , it is necessary to eliminate data heterogeneity. This can be done by centering the mean in zero and the variance in 1, for each variable from \mathbf{X} . Equation 1 is used for this purpose:

$$Z_{i,j} = \sigma_j^{-1/2}(X_{i,j} - \mu_j); i = 1, n; j = i, m, \quad (1)$$

where $X_{i,j}$ is the i th observation of variable j , $Z_{i,j}$ is the standardized variable of $X_{i,j}$, and μ_j and σ_j are the mean value and the variance of the variable j , respectively. Both n and m are the total number of observations and variables, respectively.

Table 2. Codification values for Nonnumerical Variables.

Factor	ANN Input Value
Sex	Male = 1; Female = -1
Pulmonary embolism	Yes = 1; No = -1
Site of thrombosis	Right inferior member = 1 Left inferior member = -1
Proximal or distal lower limb thrombosis	Proximal = 1; Distal = -1
Provoked or unprovoked	Provoked = 1; Unprovoked = -1
Anticoagulant use	Yes = 1; No = -1
FV Leiden	Negative = 1; Heterozygous = -1
G20210A prothrombin gene mutation	Negative = 1; Heterozygous = -1
Antiphospholipid syndrome	Negative = 1; Positive = -1
Smoking, hormonal therapy, arterial hypertension, diabetes mellitus, dyslipidemia, cancer, residual venous thrombus on US Doppler	Yes = 1; No = -1

Abbreviations: FV, factor V; US, ultrasound.

Once the matrix \mathbf{Z} is obtained, the variance-covariance matrix can be obtained. The PCs can be generated by calculating the eigenvector and the eigenvalues for the matrix \mathbf{Z} and are a linear combination of the original standardized variable:

$$PC_k = \sum_{j=1}^m a_{kj} Z_j; k = 1, m. \quad (2)$$

The importance of each PC can be determined by ordering the eigenvalues in a descending order, corresponding to the descending order of variance.

In this study, PCA was used to determine the main factors for RVTE. We considered only the PCs in which eigenvalue is greater than 1, which assures that the set of PCs are more important than the original variables. To compute the importance of each factor Z_j in a PC_k , the correlation, $\rho_{Z_j PC_k}$, was calculated by using Equation 3:

$$\rho_{Z_j PC_k} = a_{jk} \sqrt{\lambda_j}; j, k = 1, 2, \dots, m, \quad (3)$$

where a_{jk} is eigenvector value of the j th standardized variable Z with respect to the k th PCs, and λ_j is the eigenvalue of the j th standardized variable Z .

Only those variables with $|\rho_{Z_j PC_k}| > 0.5$ were considered as important factor.

Results

From all 261 patients, 26 were excluded because they presented missing data, a first VTE in other sites than pulmonary, lower limbs or central nervous system, and age under 18 years. Therefore, 235 patients were included whereas 74 (31.4%) were men and 161 were women, age >18 and <82 years.

Table 3 summarizes the principal characteristics of the thrombotic episodes. In all, 49 (20.8%) patients presented

Table 3. Main Characteristics of the Thrombotic Events.

Time between thrombotic episodes	23.75 ± 19.12 months	
Time between the end of anticoagulation therapy and thrombotic episodes	15.87 ± 14.89 months	
	VTE	RVTE
Men	74	20
Women	161	29
Location		
Left leg	107	23
Right leg	71	15
Pulmonary embolism (exclusive)	42	9
Pulmonary embolism	85	11
Central nervous system	5	1
Both legs	6	1

Abbreviations: RVTE, recurrent venous thromboembolism; VTE, venous thromboembolism.

recurrent VTE. The first episode was more predominant in the left leg (45.7%); 36% of the patients presented pulmonary embolism (PE), and only 2.1% presented thrombosis in the central nervous system. When analyzing RVTE events, the percentages were almost the same: 46.0%, 30.0%, and 2.0% for patients who presented RVTE in the left, right, and both legs, respectively. In all, 22% presented PE as RVTE as well as 2.0% in the central nervous system. Of all patients, 39% presented a provoked first thrombosis, and 40% of these patients had RVTE, corresponding to 16% of all recurrences. This is an interesting result, since recurrence in patients with a provoked VTE is expected to be low. Cancer was presented in 13 patients. Three of them had provoked VTE, and only 1 presented RVTE.

The RVTE events were not in the same location for all patients. Table 4 shows the relation the VTE/RVTE location for all patients who presented RVTE. It can be seen that the thrombotic events repeat more frequently in the same location for patients who presented VTE in the left leg: Of 24 VTE patients who presented VTE in the left leg, 16 presented RVTE in the same location. For other events, this observation is not the same.

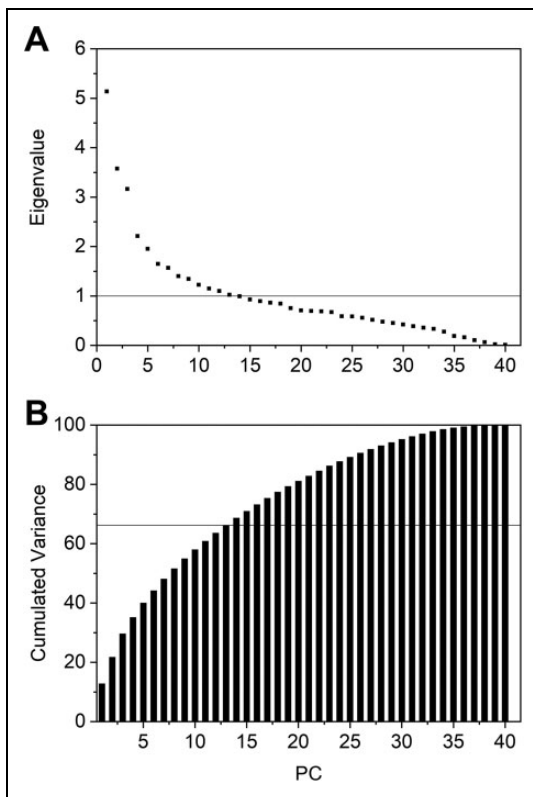
In this study, we used all gathered data to calculate the eigenvalues and the eigenvectors required to obtain the PCs, and then using Equation 3, to determine the main factors for recurrent VTE. Since we considered 39 factors, 39 PCs were generated. The eigenvalues obtained and the cumulated variance for each PC is shown in Figure 1.

The results indicated that 26 of the 39 PCs correspond to 90.37% of the overall variance. In Figure A, the line $y = 1$ corresponds to the marginal value below which a PC accounts for less variance than the original ones. So, only the PCs with eigenvalues >1 were considered. Thirteen PCs satisfy this condition, so we believe that they are sufficient to describe RVTE, corresponding to 67.34% of the total variance (horizontal line in Figure 1B).

Table 4. Relation Between VTE and RVTE Events.

VTE	RVTE						Total RVTE
	Left Leg	Right Leg	Both Legs	Central Nervous System	PE		
Left leg	16	4	0	2	1	23	
Right leg	6	6	1	2	0	15	
Both legs	0	0	1	0	0	1	
Central nervous system	0	1	0	0	0	1	
PE	2	3	1	1	2	9	
Total VTE	24	14	3	5	3	49	

Abbreviations: PE, pulmonary embolism; RVTE, recurrent venous thromboembolism; VTE, venous thromboembolism.

**Figure 1.** Principal component analysis (PCA) results. A, Eigenvalues. B, Cumulated variance.

The correlations calculated by Equation 3 between the factors and the 13 PCs considered are presented in the Supplementary Information (Table S1). A factor can be considered highly correlated with a PC, therefore an important factor, when its correlation modulus is higher than 0.5. The correlations for the other PCs (13-39) were not presented, since they were discarded. The main factors for RVTE found in this work are presented in Table 5 in order of importance.

Principal component 1 accounted for 13% of overall variance, including red blood cell (RBC) count, hematocrit (HCT), triglycerides, glucose, creatinine, cholesterol, high-density lipoprotein (HDL), and low-density lipoprotein (LDL). The former 3 factors presented much higher correlations with PC1, indicating that they have strong influence in this PC.

Table 5. Main Factors for RVTE Determined by PCA.

Principal Components		Correlation	
1	Total cholesterol	0.765	
	LDL	0.706	
	HDL	0.658	
	Creatinine	0.564	
	Triglycerides	0.551	
	Glucose	0.520	
	HCT	0.520	
	RBC	0.508	
	2	AT	0.712
		PS	0.688
		PC	0.661
	3	Age	-0.558
		HB	0.781
HCT		0.770	
5	RBC	0.749	
	WBC	0.569	
7	RDW	0.525	
	D-dimer	0.512	
9	Location of VTE in leg	0.657	
	Anticoagulation time	0.522	
12			

Abbreviations: HB, hemoglobin; HCT, hematocrit; HDL, high-density lipoprotein; LDL, low-density lipoprotein; PC, principal component; PCA, principal component analysis; RBC, red blood cell; RDW, red blood cell distribution width; RVTE, recurrent venous thromboembolism; VTE, venous thromboembolism; WBC, white blood cell; AT, Antithrombin; PS, Protein S.

Principal component 2 accounted for 9% of the overall variance, including age, protein C, protein S, and antithrombin levels. The third component accounted for 8% of the overall variance, including RBC, hemoglobin (HB), and HCT.

The other highly correlated variables were white blood cell (WBC) and red cell distribution width (RDW; PC-5: 5.7% of overall variance), D-dimer (PC-7: 4% of overall variance), proximal or distal lower limb thrombosis (PC-9: 3.5% of overall variance), and duration of anticoagulant treatment (PC-12: 2.92% of overall variance). The other PCs presented no correlation modulus higher than 0.5, indicating that there is no relative importance between the factors in those PCs.

Discussion

Based on the results, one can relate each PC with a determined cluster of variables. As we see, PC-1 is characterized by RBC,

total cholesterol, LDL and HDL cholesterol, triglycerides, glucose level, and creatinine level. Principal component 2 relates with age and natural anticoagulants levels. The main factors in PC-3 are RBC parameters. The size distribution of RBC and WBC count are the main factors in PC-5. The other PCs presented only 1 factor as important, so they are mainly characterized by them.

It is well-known that the factors for RVTE identified by PCA in this work are related to thrombotic events^{3,13,20} as transient or persistent risk factors such as recent surgery, major trauma, pregnancy, puerperium, hormone replacement therapy, and oral contraceptive use and cancer.²¹ However, since RVTE is a multifactorial disease, the exact mechanism and interactions between these factors still need a better elucidation.

Principal Component 1

The main factors given by PC-1 corroborate recent publications regarding the association of RBC, lipids, glucose, and creatinine levels with recurrent VTE. From Table S1, one can see that cholesterol, HDL, and LDL had higher correlation modulus than the others. This result means that they highly influence PC-1.

The role of the lipids in VTE is still uncertain, and contradictory results were reported.^{22,23} It was shown that HDL influences the extrinsic coagulation pathway, the protein C cascade, the fibrinolysis, and also reduces blood viscosity. Therefore, it has antithrombotic properties. On the other hand, triglycerides have procoagulant effect, suppressing the tissue plasminogen activator activity and increasing plasminogen activator inhibitor-1.²⁴ A recent prospective study²⁵ showed that there was no association between these lipid level and recurrent VTE. Eichinger et al²⁶ reported that lower levels of HDL were associated with recurrent VTE. In both studies, blood samples were collected after 3 months of discontinuation of anticoagulation. Our results showed that patients with recurrent VTE presented higher levels of total cholesterol (193.29 ± 44.25 mg/dL vs 184.90 ± 49.19 mg/dL), HDL (49.29 ± 14.91 mg/dL vs 46.41 ± 14.45 mg/dL), and LDL (114.38 ± 35.49 mg/dL vs 105.56 ± 35.04 mg/dL) than those without recurrence. On the other hand, the level of triglycerides was lower (148.52 ± 60.00 mg/dL vs 156.04 ± 152.66 mg/dL). It is important to mention that when looking at the median, it can be observed that the values for HDL are 45 mg/mL and 46 mg/mL for patients with and without RVTE, respectively. Moreover, for triglycerides, the median is 137 mg/dL and 122 mg/dL for patients with and without RVTE, respectively. Both results indicate that the statistical distribution of these factor values is in agreement with those reported in the literature.

The RBC gained importance in thrombotic events only in the last decades, when its effects in blood rheology and endothelium interactions were taken into account as a venous thrombotic risk factor.²⁷⁻²⁹ A recent review³⁰ showed that the RBC influence on recurrent thrombosis can be difficult to interpret but that antithrombotic RBC targets can be developed. Yu et al³¹ and Marchioli et al³² stated that high levels of RBC is a

potential risk factor for thrombotic events. Another proposed mechanism suggest that RBC can interact with the vessel walls, enhance platelet aggregation and activation, contribute to thrombin generation, and bind with fibrinogen contributing to thrombus size.³⁰ Our results showed that patients who presented recurrent VTE showed higher number of RBC ($4.89 \times 10^6 \pm 0.62 \times 10^6/\mu\text{L}$) than those without recurrence ($4.61 \times 10^6 \pm 0.62 \times 10^6/\mu\text{L}$), which is in agreement with cited studies.

Another factor of blood viscosity is the HCT. It was reported that increase in HCT levels in the venous system can decrease the blood flow due to the increase in its viscosity, thus favoring clot formation,^{33,34} platelet adhesion, and accumulation at the subendothelial cells.³⁰ Indeed, higher HCT levels were previously correlated with thrombotic events.²⁰ As far as we know, there is only 1 study that reported HCT level as a predictor of recurrent VTE. The study of Eischer et al³⁵ showed that patients with RVTE presented higher HCT levels (3 months after discontinuation of anticoagulation) than those without recurrence. However, higher levels of HCT correlated with recurrent VTE in women and not in men. Our results showed that HCT level in patients with RVTE were higher ($42.16\% \pm 4.39\%$) than those without ($40.70\% \pm 4.95\%$).

The role of glucose level in RVTE is still uncertain and as far as we know, there are no studies that analyzed its effects on it. High glucose levels can trigger the coagulation system in healthy men and in patients with diabetes mellitus.^{36,37} For this reason, we choose to include this factor in the study. Our results demonstrated that patients with RVTE presented lower glucose levels than those without recurrent episodes (93.12 ± 21.27 mg/dL vs 105.23 ± 51.17 mg/dL, respectively). Besides, this is not an expected result, recurrence is multifactorial, and the interaction with other factors could influence the thrombus formation.

The role of creatinine level on RVTE is still unknown. Shlipak et al³⁸ reported that higher creatinine levels are accompanied with the increase in factor VII, factor VIII, and D-dimer, which can lead to VTE events. Our results showed that patients with higher levels of creatinine have RVTE episodes against those with lower creatinine levels (0.86 ± 0.23 mg/dL vs 0.83 ± 0.31 mg/dL).

Principal Component 2

Age and coagulation proteins are the main identifiers of PC-2, and, as shown Table S1, their importance on the PC is similar. Age is a well-known important factor for thrombosis. Several research indicated that the thrombosis affects 0.01% of the population before the age of 40, and 0.7% of the population between the ages 45 and 55. In addition, the morbidity is higher as age increases.^{4,39,40} Hansson et al¹³ and Farzamia et al³ showed that age has no direct relation between age and a higher risk of recurrence. White et al⁴¹ reported that recurrence was more common in younger patients, while Beyth et al⁴² found that this was observed in patients aged 65 and younger. On the other hand, Heit et al,² Eichinger et al,²⁶ and Galanaud et al,⁴³

reported that increasing age was related to a higher risk of recurrence. Our results showed that patients with RVTE presented similar average age (43.40 ± 14.82 years) against (44.05 ± 16.07 years) those without recurrence.

Deficiency of natural anticoagulants can trigger hypercoagulability and a thrombotic event.^{36,44-47} Rosendaal and Reitsma⁴⁷ reported that these deficiencies are a strong risk factor for VTE. However, the authors affirm that this is not necessarily true for RVTE. De Stefano et al⁴⁸ indicated that patients with these deficiencies have an increased risk of recurrent VTE. Our results showed that antithrombin, protein C, and protein S activity were important factors, as patients with RVTE presented lower levels of these proteins when compared to those without recurrence (Protein C: 106.26 ± 28.39 mg-dL⁻¹; Protein S: 90.69 ± 23.83 mg-dL⁻¹; antithrombin: 105.69 ± 16.11 mg-dL⁻¹ vs Protein C: 117.87 ± 23.60 mg-dL⁻¹; Protein S: 92.41 ± 23.44 mg-dL⁻¹; antithrombin: 106.22 ± 15.57 mg-dL⁻¹, respectively). These results are very interesting as even without a classical deficiency, we showed that lower levels are important for RVTE. Our findings are in agreement Xu et al⁴⁹ who proposed a mathematical model for the coagulation cascade and showed that lower levels of protein C is associated with thrombus formation.

Principal Component 3

Principal component 3 is characterized by the RBC parameters: RBC, HB, and HCT. Both RBC and HCT were also important factors in PC-1. Since each PC is calculated as a function of all original factors, one factor can be important in more than one PC due the multidimensionality of the problem. However, their importance in PC-3 is higher. As we previously discussed, RBC and HCT can be associated with VTE and RVTE. Brækkan et al²⁰ found a direct relation between HB at the admission and VTE: The higher the HB levels, the higher the risk of VTE. Our results showed that patients with recurrence presented similar levels of HB (13.99 ± 1.44 g/dL) when compared to those without (13.54 ± 1.77 g/dL). As far as we know, there is no study that evaluated the influence of HB levels in RVTE. However, it is known that HB molecules released from damaged RBCs enhance platelet activation and aggregation.³⁰

Principal Component 5

Both RDW and WBC are the major descriptors of PC-5. The leukocyte number and the RDW have been related to thrombotic events.⁵⁰⁻⁵² However, their role in RVTE is still not well established. Some studies showed that high WBC count is related to RVTE.⁵³⁻⁵⁵ On the other hand, the study of Rezende et al⁵⁶ reported that those with higher risk of VTE presented lower levels of WBC and higher levels of RDW after VTE. The authors also suggested that other studies should be performed to investigate the relationship between these variables with RVTE. Our results showed that patients with VTE recurrence presented lower levels of WBC ($6.89 \cdot 10^3 \pm 2.69 \cdot 10^3/\mu\text{L}$) and higher levels of

RDW ($14.17 \pm 1.96\%$) than those without recurrence (WBC: $7.94 \cdot 10^3 \pm 3.86 \cdot 10^3/\mu\text{L}$; RDW: $13.59 \pm 1.49\%$).

Remaining PCs

D-Dimer, proximal or distal lower limb thrombosis, and duration of anticoagulant treatment are the descriptors of PC-7, PC-9, and PC-12, respectively. D-Dimer is the most important predictor factor for VTE and its recurrence, and several studies about its influence can be found in the literature.⁵⁷⁻⁶⁰ Eichinger et al⁶¹ and Palareti et al⁶² found that there is an association of high level of D-dimer and high recurrence risk. Also, in a recent study, Bjøri et al⁶³ reported that those with D-dimer level <1500 ng-mL⁻¹ at diagnosis were associated with a low recurrence risk. Our results showed that D-dimer level in patients with recurrence are in agreement with all the studies cited (with RVTE: 839.13 ± 1101.97 ng-mL⁻¹; without RVTE: 672.68 ± 981.58 ng-mL⁻¹).

According to Galanaud et al,⁴³ the influence of the location of VTE on the leg for RVTE is important. The authors concluded that those with proximal VTE presented more recurrent thrombotic events than those with distal VTE. Boutitie et al⁶⁴ and Hansson et al¹³ reported similar results. Our results showed that 83% of patients with distal VTE presented recurrence against 16% of those with proximal VTE.

Time of anticoagulation therapy is one of the major factors that can lead or prevent RVTE.⁶⁵ This is a subject of debate, and its recommendations can be found in the American College of Chest Physicians Guidelines.⁶⁶ Boutitie et al⁶⁴ reported that patients treated for 6 months or more presented lower RVTE events than those treated during 3 months. Also, that patients treated for 1 or 1.5 months presented higher risk of RVTE than those treated for 3 months. In a recent study, Agnelli et al¹¹ evaluated the extended anticoagulation therapy with apixaban and reported that the extended therapy reduced the risk of recurrence without increasing the risk of bleeding. Similar results are reported in Investigators.¹² Studies that analyze the time of anticoagulant after a first recurrence can also be found in the literature.^{67,68} Our results showed that patients with RVTE were treated during a longer time than those without recurrence (12.72 ± 18.51 months vs 9.88 ± 11.50 months). This result was somewhat unexpected, since patients with less time of anticoagulation therapy is supposed to have higher probability of RVTE. However, when comparing only patients with RVTE (data not shown), it could be observed that RVTE occurred mainly in patients treated during shorter periods.

The results obtained in this work showed that RVTE, a priori, can be predicted using clinical parameters and data from the previous thrombosis. It is important to mention that some variables were expected to be within the main factors, but they were not indicated by PCA, since only variables with correlation above 0.500 usually are considered as important.

The first one is the provoked/unprovoked VTE factor. Our results showed that 16% of all patients with RVTE had a first provoked VTE. On a daily basis, patients with provoked VTE are usually treated during 3 to 6 months, and then the

anticoagulation therapy is ceased. However, a different decision could avoid such a high recurrence rate. This factor is reported in Supplementary Table S1 with a correlation of 0.493 in PC-4, which is very close to 0.500. Due to its importance to RVTE, this factor could be included as input in a prediction model. Cancer is another variable expected to be in the main factors list. However, only 13 patients of 235 presented this disease. Therefore, its influence is too small when compared to the other 38 variables. This can be confirmed by its correlation coefficients in Supplementary Table S1, which are smaller than 0.360. A future study should address only these patients.

Main Limitations

The limitations for this study are mainly due the sample size, which is small considering the eligible population. However, for determination of the main factors using PCA, it was considered to be sufficient. The results showed that the important predictors of RVTE are in agreement with previous studies and are simple to obtain in clinical routine. Future work should consider another population for validation purposes.

Conclusion

In this work, PCA was applied in order to identify the main factors that predict recurrent deep vein thrombosis. Our results showed that 13 PCs are sufficient to describe the phenomenon and that 18 of 39 factors are the important for RVTE analysis. The important factors include blood cells, lipids, coagulation proteins levels, renal disease, age, as well as previous VTE parameters. This work showed that simple parameters, obtained directly from blood measurements, could be used to develop a model to predict the rates of RVTE.

Authors' Note

Each author contributed to the development of the manuscript, reviewed, and commented on each draft, and approved the final draft. R. Maciel-Filho and J. M. Annichino-Bizzacchi contributed to the study protocol. J. M. Annichino-Bizzacchi and A.V.C. Romano contributed to data acquisition. T. D. Martins and R. Maciel-Filho contributed to study design. T. D. Martins, J. M. Annichino-Bizzacchi, A.V.C. Romano, and R. Maciel-Filho contributed to data interpretation and discussion. Ethical approval to report this case was obtained from Comitê de Ética da Universidade Estadual de Campinas (Approval number: 88970218.0.0000.5404). Informed consent for patient information to be published in this article was not obtained because this was a retrospective study. No new blood samples were taken and requesting the presence of all patients would make the study unfeasible.

Acknowledgments

The authors thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), which funded this study. Process number: 2016/14172-6.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Process number: 2016/14172-6.

ORCID iD

Tiago D. Martins  <https://orcid.org/0000-0002-3452-7033>

Supplemental Material

Supplemental material for this article is available online.

References

- Hull R, Delmore T, Genton E, et al. Warfarin sodium versus low-dose heparin in the long-term treatment of venous thrombosis. *N Engl J Med.* 1979;301(16):855-858. doi:10.1056/nejm197910183011602. PubMed PMID: 384248.
- Heit JA, Mohr DN, Silverstein MD, et al. Predictors of recurrence after deep vein thrombosis and pulmonary embolism: a population-based cohort study. *Arch Int Med.* 2000;160(6):761-768. doi:10.1001/archinte.160.6.761.
- Farzamnia H, Rabiei K, Sadeghi M, et al. The predictive factors of recurrent deep vein thrombosis. *ARYA Atheroscl.* 2011;7(3):123-128. PubMed PMID: PMC3347857.
- Fahrni J, Husmann M, Gretener SB, Keo HH. Assessing the risk of recurrent venous thromboembolism—a practical approach. *Vasc Health Risk Manag.* 2015;11:451-459. doi:10.2147/VHRM.S83718. PubMed PMID: PMC4544622.
- Guyatt GH, Akl EA, Crowther M, et al. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest.* 2012;141(2 suppl):7S-47S. doi:10.1378/chest.1412S3. PubMed PMID: PMC3278060.
- Tosetto A, Iorio A, Marcucci M, et al. Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). *J Throm Haemost.* 2012;10(6):1019-1025. doi:10.1111/j.1538-7836.2012.04735.x.
- Eichinger S, Heinze G, Jandek LM, et al. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: The Vienna Prediction Model. *Circulation.* 2010;121(14):1630-1636. doi:10.1161/circulationaha.109.925214.
- Rodger MA, Kahn SR, Wells PS, et al. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. *CMAJ.* 2008;179(5):417-426. doi:10.1503/cmaj.080493. PubMed PMID: PMC2518177.
- Ensor J, Riley RD, Moore D, et al. Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE. *BMJ Open.* 2016;6(5). doi:10.1136/bmjopen-2016-011190.

10. Kyrle PA, Eichinger S. Clinical scores to predict recurrence risk of venous thromboembolism. *Thromb Haemostasis*. 2012;108(6):1061.
11. Agnelli G, Buller HR, Cohen A, et al. Apixaban for extended treatment of venous thromboembolism. *N Engl J Med*. 2013;368(8):699-708. doi:10.1056/NEJMoa1207541. PubMed PMID: 23216615.
12. Investigators TE. Oral rivaroxaban for symptomatic venous thromboembolism. *N Engl J Med*. 2010;363(26):2499-2510. doi:10.1056/NEJMoa1007903. PubMed PMID: 21128814.
13. Hansson P, Sörbo J, Eriksson H. Recurrent venous thromboembolism after deep vein thrombosis: incidence and risk factors. *Arch Int Med*. 2000;160(6):769-774. doi:10.1001/archinte.160.6.769.
14. Cosmi B, Legnani C, Cini M, et al. D-dimer and residual vein obstruction as risk factors for recurrence during and after anticoagulation withdrawal in patients with a first episode of provoked deep-vein thrombosis. *Thromb Haemostasis*. 2011;105(5):837.
15. Kutcher ME, Ferguson AR, Cohen MJ. A principal component analysis of coagulation after trauma. *J Trauma Acute Care Surg*. 2013;74(5):1223-1230. doi:10.1097/TA.0b013e31828b7fa1. PubMed PMID: PMC3638007.
16. Okin PM, Devereux RB, Fabsitz RR, et al. Principal component analysis of the T wave and prediction of cardiovascular mortality in American Indians. *Circulation*. 2002;105(6):714.
17. Thorpe MG, Milte CM, Crawford D, et al. A comparison of the dietary patterns derived by principal component analysis and cluster analysis in older Australians. *Int J Behav Nut Phys Activ*. 2016;13(1):30.
18. Vavougiou GD, Natsios G, Pastaka C, et al. Phenotypes of comorbidity in OSAS patients: combining categorical principal component analysis with cluster analysis. *J Sleep Res*. 2016;25(1):31-38. doi:10.1111/jsr.12344.
19. Chin TL, Moore EE, Moore HB, et al. A principal component analysis of postinjury viscoelastic assays: clotting factor depletion versus fibrinolysis. *Surgery*. 2014;156(3):570-577. doi:10.1016/j.surg.2014.04.030.
20. Brækkan SK, Mathiesen EB, Njølstad I, et al. Hematocrit and risk of venous thromboembolism in a general population. The Tromsø study. *Haematologica*. 2010;95(2):270-275. doi:10.3324/haematol.2009.008417. PubMed PMID: PMC2817030.
21. Heit JA. Predicting the risk of venous thromboembolism recurrence. *Am J Hematol*. 2012;87(suppl 1): S63-S67. doi:10.1002/ajh.23128. PubMed PMID: PMC3383031.
22. Deguchi H, Pecheniuk NM, Elias DJ, et al. High-density lipoprotein deficiency and dyslipoproteinemia associated with venous thrombosis in men. *Circulation*. 2005;112(6):893.
23. Everett BM, Glynn RJ, Buring JE, et al. Lipid biomarkers, hormone therapy, and the risk of venous thromboembolism in women. *JTH*. 2009;7(4):588-596. doi:10.1111/j.1538-7836.2009.03302.x. PubMed PMID: PMC2694669.
24. Belaj K, Hackl G, Rief P, et al. Changes in lipid metabolism and extension of venous thromboembolism. *Ann Nut Metabol*. 2014;64(2):122-126.
25. Morelli VM, Lijfering WM, Rosendaal FR, et al. Lipid levels and risk of recurrent venous thrombosis: results from the MEGA follow-up study. *J Throm Haemost*. 2017;15(4):695-701. doi:10.1111/jth.13640.
26. Eichinger S, Pecheniuk NM, Hron G, et al. High-density lipoprotein and the risk of recurrent venous thromboembolism. *Circulation*. 2007;115(12):1609.
27. Litvinov RI, Weisel JW. Role of red blood cells in haemostasis and thrombosis. *ISBT Sci Series*. 2017;12(1):176-183. doi:10.1111/voxs.12331.
28. Alt E, Banyai S, Banyai M, et al. Blood rheology in deep venous thrombosis—relation to persistent and transient risk factors. *Thromb Res*. 2002;107(3):101-107. doi:10.1016/S0049-3848(02)00302-X.
29. Vayá A, Suescun M. Hemorheological parameters as independent predictors of venous thromboembolism. *Clin Hemorheol Microcirc*. 2013;53(1-2):131-141.
30. Byrnes JR, Wolberg AS. Red blood cells in thrombosis. *Blood*. 2017;130(16):1795-1799. doi:10.1182/blood-2017-03-745349.
31. Yu FTH, Armstrong JK, Tripette J, et al. A local increase in red blood cell aggregation can trigger deep vein thrombosis: evidence based on quantitative cellular ultrasound imaging. *J Throm Haemos*. 2011;9(3):481-488. doi:10.1111/j.1538-7836.2010.04164.x.
32. Marchioli R, Finazzi G, Specchia G, et al. Cardiovascular events and intensity of treatment in polycythemia vera. *N Engl J Med*. 2013;368(1):22-33. doi:10.1056/NEJMoa1208500. PubMed PMID: 23216616.
33. Wells RE, Merrill EW. Influence of flow properties of blood upon viscosity-hematocrit relationships. *J Clin Invest*. 1962;41(8):1591-1598. PubMed PMID: PMC291077.
34. Dintenfass L. Viscosity and clotting of blood in venous thrombosis and coronary occlusions. *Circ Res*. 1964;14(1):1.
35. Eischer L, Tscholl V, Heinze G, et al. Hematocrit and the risk of recurrent venous thrombosis: a prospective cohort study. *PLoS One*. 2012;7(6):e38705. doi:10.1371/journal.pone.0038705.
36. Kim HK, Kim JE, Park SH, et al. High coagulation factor levels and low protein C levels contribute to enhanced thrombin generation in patients with diabetes who do not have macrovascular complications. *J Diabetes Complications*. 2014;28(3):365-369. doi:10.1016/j.jdiacomp.2014.01.006.
37. Grant PJ. Diabetes mellitus as a prothrombotic condition. *J Int Med*. 2007;262(2):157-172. doi:10.1111/j.1365-2796.2007.01824.x.
38. Shlipak MG, Fried LF, Crump C, et al. Elevations of inflammatory and procoagulant biomarkers in elderly persons with renal insufficiency. *Circulation*. 2003;107(1):87-92. doi:10.1161/01.cir.0000042700.48769.59.
39. Tsai AW, Cushman M, Rosamond WD, et al. Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology. *Arch Int Med*. 2002;162(10):1182-1189. doi:10.1001/archinte.162.10.1182.
40. Silverstein MD, Heit JA, Mohr DN, et al. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Int Med*. 1998;158(6):585-593. doi:10.1001/archinte.158.6.585.
41. White RH, Zhou H, Romano PS. Length of hospital stay for treatment of deep venous thrombosis and the incidence of

- recurrent thromboembolism. *Arch Int Med*. 1998;158(9): 1005-1010. doi:10.1001/archinte.158.9.1005.
42. Beyth RJ, Cohen AM, Landefeld C. Long-term outcomes of deep-vein thrombosis. *Arch Int Med*. 1995;155(10):1031-1037. doi:10.1001/archinte.1995.00430100053006.
43. Galanaud JP, Sevestre MA, Genty C, et al. Incidence and predictors of venous thromboembolism recurrence after a first isolated distal deep vein thrombosis. *J Thrombos Haemost*. 2014;12(4): 436-443. doi:10.1111/jth.12512.
44. Heeb MJ, Rosing J, Bakker HM, et al. Protein S binds to and inhibits factor Xa. *Proceed Nat Acad Sci*. 1994;91(7):2728-2732.
45. Pabinger I, Schneider B. Thrombotic risk in hereditary antithrombin III, protein C, or protein S deficiency. *Cooperat Retrospect Stud*. 1996;16(6):742-748. doi:10.1161/01.atv.16.6.742.
46. Esmon C. The protein C pathway. *Crit Care Med*. 2000;28(9): S44-S8. PubMed PMID: 00003246-200009001-00010.
47. Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. *J Thrombos Haemost*. 2009;7:301-304. doi:10.1111/j.1538-7836.2009.03394.x.
48. De Stefano V, Simioni P, Rossi E, et al. The risk of recurrent venous thromboembolism in patients with inherited deficiency of natural anticoagulants antithrombin, protein C and protein S. *Haematologica*. 2006;91(5):695.
49. Xu Z, Lioi J, Mu J, et al. A multiscale model of venous thrombus formation with surface-mediated control of blood coagulation cascade. *Biophys J*. 2010;98(9):1723-1732. doi:10.1016/j.bpj.2009.12.4331. PubMed PMID: PMC2862154.
50. LoPresti R, Ferrara F, Canino B, et al. Deep venous thrombosis: leukocyte rheology at baseline and after in vitro activation. *Pathophysiol Haemos Throm*. 2000;30(4):168-173.
51. Gangat N, Strand J, Li CY, et al. Leucocytosis in polycythaemia vera predicts both inferior survival and leukaemic transformation. *Brit J Haematol*. 2007;138(3):354-358. doi:10.1111/j.1365-2141.2007.06674.x.
52. Bucciarelli P, Maino A, Felicetta I, et al. Association between red cell distribution width and risk of venous thromboembolism. *Thromb Res*. 2015;136(3):590-594. doi:10.1016/j.thromres.2015.07.020.
53. De Stefano V, Za T, Rossi E, et al. Recurrent thrombosis in patients with polycythemia vera and essential thrombocythemia: incidence, risk factors, and effect of treatments. *Haematologica*. 2008;93(3):372-380. doi:10.3324/haematol.12053.
54. Saraiva SDS, Custódio IF, Mazetto BDM, et al. Recurrent thrombosis in antiphospholipid syndrome may be associated with cardiovascular risk factors and inflammatory response. *Thromb Res*. 2015;136(6):1174-1178. doi:10.1016/j.thromres.2015.10.029.
55. Carobbio A, Finazzi G, Guerini V, et al. Leukocytosis is a risk factor for thrombosis in essential thrombocythemia: interaction with treatment, standard risk factors, and Jak2 mutation status. *Blood*. 2007;109(6):2310-2313. doi:10.1182/blood-2006-09-046342.
56. Rezende SM, Lijfering WM, Rosendaal FR, et al. Hematological variables and venous thrombosis: red cell distribution width and blood monocytes are associated with an increased risk. *Haematologica*. 2014;99(1):194-200.
57. Lippi G, Favaloro EJ, Cervellin G, eds. A review of the value of D-dimer testing for prediction of recurrent venous thromboembolism with increasing age. *Seminars in Thrombosis and Hemostasis*. New York, NY: Thieme Medical Publishers; 2014.
58. Palareti G, Legnani C, Cosmi B, et al. Predictive value of D-dimer test for recurrent venous thromboembolism after anticoagulation withdrawal in subjects with a previous idiopathic event and in carriers of congenital thrombophilia. *Circulation*. 2003;108(3): 313-318. doi:10.1161/01.cir.0000079162.69615.0 f.
59. Verhovsek M, Douketis JD, Yi Q, et al. Systematic review: D-dimer to predict recurrent disease after stopping anticoagulant therapy for unprovoked venous thromboembolism. *Ann Intern Med*. 2008;149(7):481-490. doi:10.7326/0003-4819-149-7-200810070-00008.
60. Palareti G, Legnani C, Cosmi B, et al. Risk of venous thromboembolism recurrence: high negative predictive value of D-dimer performed after oral anticoagulation is stopped. *Thromb Haemost*. 2002;87(1):7-12.
61. Eichinger S, Minar E, Bialonczyk C, et al. D-dimer levels and risk of recurrent venous thromboembolism. *JAMA*. 2003;290(8): 1071-1074. doi:10.1001/jama.290.8.1071.
62. Palareti G, Cosmi B, Legnani C, et al. D-dimer to guide the duration of anticoagulation in patients with venous thromboembolism: a management study. *Blood*. 2014;124(2):196-203. doi:10.1182/blood-2014-01-548065.
63. Bjøri E, Johnsen HS, Hansen JB, et al. D-dimer at venous thrombosis diagnosis is associated with risk of recurrence. *J Thromb Haemost*. 2017;15(5):917-924. doi:10.1111/jth.13648.
64. Boutitie F, Pinede L, Schulman S, et al. Influence of preceding length of anticoagulant treatment and initial presentation of venous thromboembolism on risk of recurrence after stopping treatment: analysis of individual participants' data from seven trials. *BMJ*. 2011;342. doi:10.1136/bmj.d3036.
65. Franco L, Giustozzi M, Agnelli G, et al. Anticoagulation in patients with isolated distal deep vein thrombosis: a meta-analysis. *J Thromb Haemos*. 2017;15(6):1142-1154. doi:10.1111/jth.13677.
66. Kearon C, Akl EA, Ornella J, et al. Antithrombotic therapy for VTE disease. *Chest*. 2016;149(2):315-352. doi:10.1016/j.chest.2015.11.026.
67. van der Hulle T, Tan M, den Exter PL, et al. Recurrence risk after anticoagulant treatment of limited duration for late, second venous thromboembolism. *Haematologica*. 2015;100(2): 188-193. doi:10.3324/haematol.2014.112896.
68. Schulman S, Granqvist S, Holmström M, et al. The duration of oral anticoagulant therapy after a second episode of venous thromboembolism. *N Engl J Med*. 1997;336(6):393-398.