ESMO Open Cancer Horizons



# Real-world features associated with cancer-related venous thromboembolic events

Maija Helena Peippo,[1] Samu Kurki,[2] Riitta Lassila,[3] Olli Mikael Carpén[4]

**To cite:** Peippo MH, Kurki S, Lassila R, *et al*. Real-world features associated with cancer-related venous thromboembolic events. *ESMO Open* 2018;**3**:e000363. doi:10.1136/esmoopen-2018-000363

Received 20 March 2018
Revised 30 May 2018
Accepted 2 June 2018



[1]Institute of Biomedicine, Research Center for Cancer, Infections and Immunity, University of Turku, Turku, Finland
[2]Auria Biobank, University of Turku and Turku University Hospital, Turku, Finland
[3]Unit of Coagulation Disorders, Department of Hematology and Comprehensive Cancer Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[4]Genome Scale Biology Research Program and Department of Pathology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

**Correspondence to**
Prof. Olli Mikael Carpén; olli.carpen@helsinki.fi

## ABSTRACT

**Background** The incidence of venous thromboembolism (VTE) is 1–2/1000 individuals. Patients with cancer, especially during chemotherapy, are at enhanced risk, but real-world data on factors associated with VTE events are still scarce.

**Aim** The aim of this retrospective study was to survey the incidence of VTE based on a large hospital database, and to identify comorbidities and features associated with VTE events. We focused on cancer-related VTE events and on factors indicating increased VTE risk during chemotherapy.

**Methods** The cohort included patients treated at Turku University Hospital during years 2005–2013. Health information was derived and analysed from multiple electronic databases. The diagnoses of VTE and all comorbidities, including type of cancer, were based on International Classification of Diseases 10th Revision coding. For further analysis, we focused on 16 common types of cancers treated with chemotherapy. Age, gender, surgery, radiotherapy, distant metastasis, available laboratory values and platinum-based chemotherapy were evaluated for VTE group, and associations were estimated by Cox regression analyses.

**Results** The entire database contained information from 495 089 patients, of whom 5452 (1.1%) had a VTE diagnosis. Among individuals with VTE, 1437 (26.4%) had diagnosis of coronary heart disease and 1467 (26.9%) had cancer diagnosis. Among 7778 patients with cancer treated with chemotherapy, 282 (3.6%) had a VTE, platinum-based chemotherapy being a major risk factor (HR 1.77, 95% CI 1.40 to 2.24, p<0.001). In multivariate analysis, elevated blood neutrophil counts (>3.25×$10^9$ cells/L, HR 1.96, 95% CI 1.33 to 2.89, p<0.001) and plasma creatinine (>62.5 µmol/L; HR 1.60, 95% CI 1.21 to 2.13, p=0.001) values were independent indicators of increased VTE risk during chemotherapy.

**Conclusions** Longitudinal electronic health record analysis provides a powerful tool to gather meaningful real-world information to study clinical associations, like comorbidities, and to identify markers associated with VTE. The combination of various clinical and laboratory variables could be used for VTE risk evaluation and targeted prevention.

## Key Questions

### What is already known about this subject?

- The incidence of venous thromboembolisms (VTE) is 1–2/1000 individuals. Patients with cancer, especially during chemotherapy and radiotherapy, are at enhanced risk of VTE. The risk of VTE depends on cancer type, and also on many additional variables, which makes prognostic evaluation challenging.

### What does this study add?

- The combination of all medically relevant longitudinal electronic data sets provides in-depth real-world information of VTE risk factors associated with cancer.

### How might this impact on clinical practice?

- The combination of various clinical and laboratory variables could form a new basis for VTE risk evaluation and targeted prevention.

## INTRODUCTION

Venous thromboembolism (VTE) is a disorder, in which blood clots are formed in deep veins (deep vein thrombosis, DVT). VTE may lead to thromboembolism pulmonary arteries (pulmonary embolism, PE) or in other organs, with or without symptoms.[1] The estimated annual incidence rate of VTE ranges from 100 to 200 among 100 000 individuals of Caucasian origin.[2] A variety of factors affect VTE risk, including major surgery, multiple traumas, inflammation and infections.[3 4] While major morbidity and mortality[5 6] are associated with VTE, it is a medical condition that can be treated or even prevented.[7–9]

Cancer is a major risk factor for VTE[4 10]; 1.6% of patients with cancer may encounter VTE.[11] The VTE risk associated with cancer is up to 10-fold as compared with general population. The pathogenesis of blood coagulation activation in cancer is complex, reciprocal and multifactorial,[12] but the epidemiological, clinical and laboratory aspects of association between cancer and VTE have provided important insights. The risk depends on cancer type[13–15] and the highest incidence of VTE has been associated with adenocarcinomas. Chemotherapy, cancer surgery and radiotherapy are all also well recognised as important risk factors for VTE,[4 15–17] as is the advanced disease stage.[18 19]

The effect of comorbidities on VTE among patients with cancer has also been evaluated.[20–22] The risk factors are often combined and synergistic, one example including surgical management of pathological fractures.[23] Many of the previous risk factor analyses suffer from lack of comprehensive clinical information, which typically limits the generalisation of the findings. Real-world data are widely needed to provide the information of clinical outcomes outside the clinical trial settings, as Kaatz *et al* observed on their study of the duration of anticoagulant treatment.[24] At the era of longitudinal electronic health records and the tools to combine large data sets, it is now possible to acquire real-world information of the factors contributing to VTE risk, to use a machine learning approach for risk assessment of VTE[25 26] and to identify markers that could be used for targeted VTE prevention.

As a first step towards these goals, we surveyed the incidence of VTE in a large hospital cohort, studied clinical factors, such as comorbidities associated with VTE, and identified cancer-associated features in VTE events. We combined data sets from various longitudinal electronic health records of Turku University Hospital, Finland. For further analysis of the predictors of VTE, we chose to focus on 16 different types of malignant disorders (advanced or metastatic) treated with chemotherapy. Age, gender, surgery, radiotherapy, distant metastasis and platinum-based chemotherapy were evaluated for VTE group, and associations were estimated by Cox regression analyses. Additionally, with the unique data source, we could also analyse the correlation of multiple routinely used laboratory variables with VTE.

## PATIENTS AND METHODS

### Data Source

The data set was created by combining several individual Turku University Hospital electronic health records, using the license of Auria Biobank, linked with Turku University Hospital Patient Discharge Data and Population Register Center. The hospital database of the clinical information is documented in a structured format, including patient data (gender, date of birth), International Classification of Diseases 10th Revision (ICD-10) codes for clinical diagnoses, structured systematic pathology reports (SNOMED) with pathological TNM for pathology, Nordic Classification of Surgical Procedures coding for medical procedures, periods of intravenous chemotherapy or radiation therapy, and routinely monitored laboratory values of clinical chemistry and haematology, which are used in the diagnosis and follow-up of the patients. All diagnoses and ICD-10 codes were provided and inserted to the electronic health records by the physicians responsible for the patient. The starting date of this survey was 1 January 2005, and the records are documented electronically in Turku University Hospital up to the last follow-up date, 31 August 2013. The Population Register Center provided the dates of death, which were used in the survival analysis.

### Cohort selection

The study population comprised only adult patients (aged 18 years and over) who were treated in Turku University Hospital either the inpatient or outpatient setting. This is a public tertiary hospital covering the region of Southwest Finland (population base of 0.5 million). The hospital covers all medical disciplines and is the only primary cancer treatment facility for this population. As search criteria for VTE, we used the ICD-10 codes for DVT (ICD-10 code I80.2), PE (ICD-10 I26) or portal vein thrombosis (ICD-10 I81). The ICD-10 codes cannot distinguish between symptomatic and asymptomatic VTE cases, and therefore both were included.

### Comorbidities and matched general population as the comparison cohort

The comorbidity was defined by main or secondary diagnosis in the electronic health records from hospital register according to ICD-10 codes that are listed in table 1. We combined the most common chronic diseases into larger categories based on vital organs and disease groups according to medical classification by ICD-10. To compare how VTE events were distributed between different disease groups, we selected age and gender-matched control population with the same diagnosis codes of chronic diseases. Thus, all patients with VTE were age and gender matched with a fivefold control population without a VTE diagnosis in their medical history.

### Covariates in the analysis in patients with cancer treated with chemotherapy

In further analysis, we focused only on patients with cancer who were treated with chemotherapy during the disease course (for neoadjuvant, adjuvant or palliative intent). Among those patients, the covariates, including laboratory values, were available due to routine follow-up visits. We focused on 16 cancer types with information from over 100 patients, who had been treated with chemotherapy during the 8-year follow-up period (2005–2013). The cancer types were breast, bladder, colorectal, gastric, Hodgkin's lymphoma, leukaemia, lung, melanoma, mesothelioma, myeloma, non-Hodgkin's lymphoma, ovarian, pancreatic, prostate, testicular and endometrial. The diagnoses were based on ICD-10 codes listed in table 2. The covariates included age (at the initiation of chemotherapy), gender, laboratory values (starting from the beginning of chemotherapy), type of surgery (either for curative or palliative intent, where tumour tissue was removed for histology), distant metastases (ICD-10 codes C78*–C79*) and platinum-based chemotherapy (for neoadjuvant, adjuvant or palliative intent).

The surgery procedure was based on the information of the available surgical sample of the tumour (block of tissue) in the pathology database. Platinum-based chemotherapy was included, because it is used as single or combination therapy in a variety of cancer types, and the information was readily aggregated in the hospital

**Table 1** Prevalence of different diseases in the patients with VTE among hospitalised patients, including both inpatient and outpatient information. For each patient with VTE diagnosis, we randomly selected five control patients (fivefold matching) with the same gender and year of birth from the Turku University Hospital electronic health records

| Disease | ICD-10 code | Number of VTE† | Per cent of patients with VTE† diagnosis | Number of control patients (fivefold matching) | Per cent of control patients (from all patients) | RR | 97.5% CI | P values‡ |
|---|---|---|---|---|---|---|---|---|
| Obesity | E65–E66* | 332 | 6.1 | 622 | 2.3 | 2.67 | 2.35 to 3.03 | <0.001 |
| Liver disease | K70*–K74* | 122 | 2.2 | 269 | 1.0 | 2.27 | 1.84 to 2.79 | <0.001 |
| Congestive heart failure | I50* | 1091 | 20.0 | 2600 | 9.5 | 2.10 | 1.97 to 2.24 | <0.001 |
| Asthma | J45*–J46 | 462 | 8.5 | 1124 | 4.1 | 2.06 | 1.85 to 2.28 | <0.001 |
| Varicose veins | I83* | 265 | 4.9 | 647 | 2.4 | 2.05 | 1.78 to 2.35 | <0.001 |
| Pulmonary diseases (not asthma) | J41*–J44*, J47, J60–J70* | 537 | 9.8 | 1366 | 5.0 | 1.97 | 1.79 to 2.16 | <0.001 |
| Rheumatoid arthritis | M05*–M06* | 298 | 5.5 | 761 | 2.8 | 1.96 | 1.72 to 2.23 | <0.001 |
| Inflammatory bowel disease | K50*–K51* | 117 | 2.1 | 302 | 1.1 | 1.94 | 1.57 to 2.39 | <0.001 |
| Peripheral vascular disease | I70*–I79* | 507 | 9.3 | 1434 | 5.3 | 1.77 | 1.61 to 1.95 | <0.001 |
| Psychiatric disease | F10*–F99 | 795 | 14.6 | 2327 | 8.5 | 1.71 | 1.58 to 1.84 | <0.001 |
| Coronary heart disease | I20*–I25* | 1437 | 26.4 | 4330 | 15.9 | 1.66 | 1.57 to 1.75 | <0.001 |
| Hypertension | I10–I15* | 2296 | 42.1 | 7405 | 27.2 | 1.55 | 1.49 to 1.61 | <0.001 |
| Cancer | C00*–C99* | 1467 | 26.9 | 4820 | 17.7 | 1.52 | 1.45 to 1.60 | <0.001 |
| Diabetes mellitus | E10*–E14* | 869 | 15.9 | 2898 | 10.6 | 1.50 | 1.40 to 1.61 | <0.001 |
| Atrial fibrillation/ flutter | I48 | 999 | 18.3 | 3881 | 14.2 | 1.29 | 1.21 to 1.37 | <0.001 |
| Cerebrovascular disease | I60*–I69* | 698 | 12.8 | 2751 | 10.1 | 1.27 | 1.17 to 1.37 | <0.001 |
| Pregnancy/delivery | O00*–O99* | 155 | 2.8 | 754 | 2.8 | 1.03 | 0.87 to 1.22 | 0.752 |

Hospitalised patients overall: 495 089.
Number of patients with VTE† overall: 5452.
Per cent of patients with VTE† diagnosis: 1.1.
Number of control patients overall: 27 260.
†VTE was defined as pulmonary embolism (PE, ICD-10 I26*), deep venous thrombosis (DVT, ICD-10 I80.2*) or portal vein thrombosis (PVT, ICD-10 I81*).
‡P values were calculated with $X^2$ test.
ICD-10, International Classification of Diseases 10th Revision; RR, relative risk; VTE, venous thromboembolism.

pharmacy database. The clinical chemistry and haematology measurements were included, if the test was performed at least 80% of patients both within 3 months before chemotherapy and during chemotherapy. Thirteen laboratory variables fulfilled the criteria, and were included for further analyses: blood cell counts and characteristics: platelet count (B-PLT), haemoglobin (B-Hgb) level, leucocyte count (B-Leuk), neutrophil count (B-Neut), erythrocyte mean cellular volume and mean corpuscular haemoglobin (E-MCV and E-MCH) and haematocrit. In addition, plasma creatinine (P-Crea), plasma alanine transaminase, plasma alkaline phosphatase, plasma sodium (P-Na) and plasma potassium (P-K) values were collected.

### Statistical analysis

All statistical analyses were performed using R Statistics V.3.0.2 with standard packages (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria). The $X^2$ test was used to calculate relative risks (RR) for comparing the proportions of patients with VTE to non-VTE patients between different groups of disease.

Age, gender, surgery, radiotherapy, distant metastases and platinum-based chemotherapy were evaluated for VTE group and the associations were estimated by Cox regression analyses in both univariate and multivariate models. The Cox proportional hazards models were used

Table 2 Incidence of VTE according to different cancer types

| Cancer | ICD-10 code | Number of patients with cancer | Number of patients with VTE† | Per cent of patients with VTE† diagnosis |
|---|---|---|---|---|
| Mesothelioma | C45* | 207 | 14 | 6.8 |
| Gastric | C16* | 898 | 54 | 6.0 |
| Ovarian | C56* | 936 | 53 | 5.7 |
| Pancreatic | C25* | 1080 | 59 | 5.5 |
| Lung | C33*–C34* | 2693 | 148 | 5.5 |
| Myeloma | C90* | 848 | 45 | 5.3 |
| Colorectal | C18*–C21* | 3727 | 168 | 4.5 |
| Non-Hodgkin's lymphoma | C82*–C85* | 1837 | 80 | 4.4 |
| Bladder | C66*–C68* | 1941 | 83 | 4.3 |
| Endometrial | C54* | 1645 | 67 | 4.1 |
| Leukaemia | C91*–C95* | 1475 | 46 | 3.1 |
| Breast | C50* | 7132 | 206 | 2.9 |
| Prostate | C61* | 7310 | 199 | 2.7 |
| Hodgkin's lymphoma | C81* | 402 | 11 | 2.7 |
| Melanoma | C43* | 1736 | 46 | 2.6 |
| Testicular | C62* | 345 | 6 | 1.7 |

Patients with cancer overall: 42 245.
Number of patients with cancer and VTE†: 1467.
Per cent of patients with VTE† diagnosis: 3.5.
†VTE was defined as pulmonary embolism (PE, ICD-10 I26*), deep venous thrombosis (DVT, ICD-10 I80.2*) or portal vein thrombosis (PVT, ICD-10 I81*).
ICD-10, International Classification of Diseases 10th Revision; VTE, venous thromboembolism.

to calculate HRs and 95% CIs. The Cox model time scale included follow-up time from the first chemotherapy treatment to the date of VTE diagnosis, or date of death or the end of follow-up (31 August 2013). P values <0.05 were considered as statically significant.

Receiver operating characteristic (ROC) curves alongside of 95% CIs were produced to explore the specific application to the choice of optimal laboratory variables on predicting of VTE. Continuous laboratory values were dichotomised into categorical variables according to the cut-off values with a maximum sum of sensitivity and specificity. The associations between VTE and the cut-off values were estimated by univariate and multivariate Cox regression models.

Kaplan-Meier analysis was used to assess the time to VTE and the VTE-free time from the beginning of chemotherapy to documented VTE event or patient's death or the end of follow-up, whichever occurred first.

## RESULTS

### Cohort characteristics and comorbidities associated with VTE

The entire electronic database contained information from 495 089 hospitalised patients (including both inpatient and outpatient), of whom 5452 (1.1%) had VTE diagnosis (table 1). The most common diagnoses associated with VTE were cardiovascular disorders (hypertension, coronary heart disease and congestive heart failure) and cancer. A simultaneous VTE and cancer diagnosis was made for 1467 patients; together, 26.9% of all patients with VTE carried a concomitant cancer diagnosis. As expected, the patients with VTE often had several concomitant diagnoses; therefore, the number of VTE and diagnosis combinations does not equal/match with the number of actual patients with VTE in table 1.

To observe which diagnoses were over-represented in the VTE population, we collected an age and gender-matched control group without any indication of a VTE event in patient records (n=27 260). Coronary heart disease (n=1437) was associated with increased incidence of VTE (RR 1.66, 95% CI 1.57 to 1.75, p<0.001). A similar association was seen with obesity (RR 2.67, 95% CI 2.35 to 3.03, p<0.001) and other cardiovascular diagnoses including atrial fibrillation/flutter, coronary heart disease, hypertension, cerebrovascular disease or congestive heart failure. An increased risk of VTE was also evident in patients with inflammatory bowel disease (RR 1.94, 95% CI 1.57 to 2.39, p<0.001). Additionally, 795 (14.6%) patients with VTE diagnosis had concurrent ICD codes for psychiatric disease (F10*–F99), indicating it as a significant risk of VTE (RR 1.71, 95% CI 1.58 to 1.84, p<0.001). Pregnancy and delivery were not associated with the increased risk of VTE (RR 1.03, 95% CI 0.87 to 1.22, p=0.752). In general, cancer was associated

**Table 3** Characteristics of chemotherapy-treated patients (n=7778) in 16 different types of malignancies

| Cancer | ICD-10 code | Patients (%) | Gender | | Age | Surgery (%) | Radiotherapy (%) | Platinum-based chemotherapy (%) | Distant metastases (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | | Female (%) | Male (%) | Median (SD) | | | | |
| Breast | C50* | 2277 (29.3) | 99.5 | 0.5 | 57.7 (10.2) | 91.9 | 80.8 | 1.8 | 12.4 |
| Bladder | C66*–C68* | 567 (7.3) | 19.0 | 81.0 | 71.1 (11.3) | 31.9 | 14.5 | 16.0 | 6.7 |
| Colorectal | C18*–C21* | 679 (8.7) | 46.2 | 53.8 | 63.6 (9.9) | 67.5 | 40.2 | 71.0 | 47.9 |
| Gastric | C16* | 203 (2.6) | 37.4 | 62.6 | 62.5 (9.7) | 52.2 | 40.9 | 47.8 | 26.6 |
| Hodgkin's lymphoma | C81* | 117 (1.5) | 51.3 | 48.7 | 37.4 (18.2) | 58.1 | 59.8 | 6.0 | 0.9 |
| Leukaemia | C91*–C95* | 422 (5.4) | 42.4 | 57.6 | 48.5 (25.2) | 8.3 | 8.8 | 2.6 | 1.2 |
| Lung | C33*–C34* | 981 (12.6) | 32.3 | 67.7 | 66.0 (9.1) | 24.5 | 51.3 | 82.6 | 15.0 |
| Melanoma | C43* | 115 (1.5) | 34.8 | 65.2 | 61.9 (12.7) | 73.9 | 54.8 | 40.0 | 43.5 |
| Mesothelioma | C45* | 110 (1.4) | 15.5 | 84.5 | 66.4 (8.5) | 25.5 | 40.0 | 95.5 | 7.3 |
| Myeloma | C90* | 276 (3.5) | 45.7 | 54.3 | 64.1 (9.0) | 10.1 | 29.0 | 2.5 | 5.1 |
| Non-Hodgkin's lymphoma | C82*–C85* | 680 (8.7) | 44.6 | 55.4 | 61.8 (14.7) | 49.1 | 41.9 | 7.6 | 5.4 |
| Ovarian | C56* | 363 (4.7) | 100 | 0 | 63.7 (12.0) | 88.2 | 14.6 | 96.4 | 10.7 |
| Pancreatic | C25* | 283 (3.6) | 47.0 | 53.0 | 63.9 (8.9) | 21.6 | 17.3 | 9.5 | 36.0 |
| Prostate | C61* | 325 (4.2) | 0 | 100 | 70.5 (7.7) | 34.2 | 83.4 | 9.5 | 43.7 |
| Testicular | C62* | 133 (1.7) | 0 | 100 | 33.5 (11.1) | 66.9 | 9.8 | 98.5 | 3.8 |
| Endometrial | C53*–54* | 247 (3.2) | 100 | 0 | 67.3 (10.1) | 89.1 | 64.8 | 88.3 | 8.1 |

Patients: 7778.
VTE†: 282 (3.6%).
Age, median (years)=61.2 (SD 13.9).
Gender: female=4549 (58.5%), male=3229 (41.5%).
Radiotherapy: 3906 (50.2%).
Platinum-based chemotherapy: 2507 (32.2%).
Surgery: 4457 (57.3%).
Distant metastasis: 1270 (16.3%).
†VTE was defined as pulmonary embolism (PE, ICD-10 I26*), deep venous thrombosis (DVT, ICD-10 I80.2*) or portal vein thrombosis (PVT, ICD-10 I81*).
ICD-10, International Classification of Diseases 10th Revision; VTE, venous thromboembolism.

with increased number of VTE events (RR 1.52, 95% CI 1.45 to 1.60, p<0.001).

### Cancer as a risk factor for VTE

The database contained information from a significant number of 42 245 patients with cancer, of whom 1467 (3.5%) had the diagnosis of VTE (table 2). Of different cancer types, the highest VTE rates were observed in mesothelioma (6.8%), followed by gastric (6.0%), ovarian (5.7%), pancreatic (5.5%) and lung cancers (5.5%) and myeloma (5.3%). The lowest VTE events were documented in melanoma (2.6%) and testicular cancer (1.7%) (table 2). Overall, patients with cancer had threefold more VTE events as compared with the entire hospital patient population (3.2% vs 1.1%).

### Risk factors for VTE in chemotherapy-treated patients with cancer

A total of 7778 patients with cancer had received chemotherapy during the disease course (either neoadjuvant, adjuvant or palliative), of whom 4549 (58.5%) were female and 3229 (41.5%) male. The median age was 61.2 years (SD 13.9). Of these patients, 284 (3.6%) were diagnosed with VTE (table 3), and the median time from the onset of chemotherapy to VTE diagnosis was 7.3 months (figure 1).

Sixteen per cent of the patients with cancer receiving chemotherapy were diagnosed with advanced-stage disease (distant metastasis according to ICD-10 codes C78*–C79*), 57% had undergone surgery (sample of

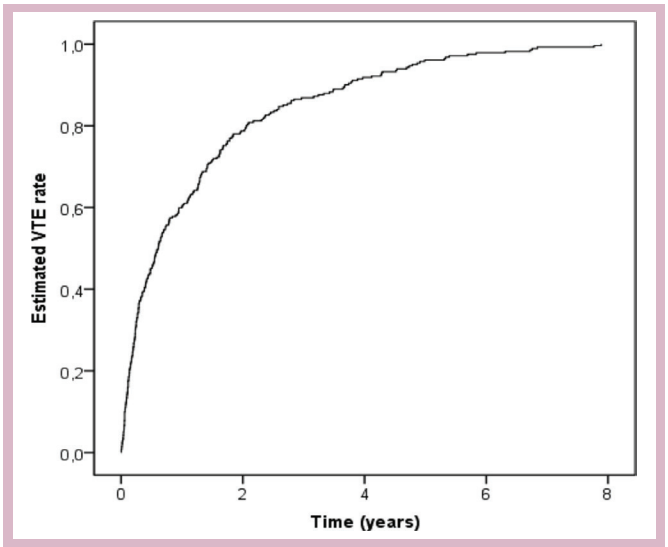


**Figure 1** Kaplan-Meier analysis of time to venous thromboembolism (VTE) from the onset of chemotherapy for patients with diagnosed VTE (n=282).

**Table 4** HRs of the association of VTE and covariates by univariate and multivariate Cox regression analysis of the patients treated with chemotherapy (n=7778)

| | Univariate | | | Multivariate | | |
|---|---|---|---|---|---|---|
| | HR | 95% CI | P values | HR | 95% CI | P values |
| Age | 1.03 | 1.02 to 1.04 | <0.001 | 1.02 | 1.01 to 1.03 | 0.0011 |
| Gender (male vs female) | 1.34 | 1.09 to 1.74 | 0.0080 | 1.13 | 0.83 to 1.53 | 0.45 |
| Distant metastasis | 2.11 | 1.62 to 2.76 | <0.001 | 1.89 | 1.40 to 2.53 | <0.001 |
| Platinum-based chemotherapy | 1.77 | 1.40 to 2.24 | <0.001 | 0.83 | 0.56 to 1.21 | 0.33 |
| Radiotherapy | 0.85 | 0.68 to 1.08 | 0.18 | 0.98 | 0.75 to 1.28 | 0.87 |
| Surgery | 0.74 | 0.59 to 0.94 | 0.014 | 1.02 | 0.76 to 1.36 | 0.91 |
| Cancer | | | | | | |
| Breast | Reference group | | | Reference group | | |
| Bladder | 1.09 | 0.58 to 2.04 | 0.80 | 0.82 | 0.39 to 1.72 | 0.60 |
| Colorectal | 2.48 | 1.57 to 3.92 | <0.001 | 1.87 | 1.03 to 3.38 | 0.038 |
| Gastric | 2.52 | 1.20 to 5.33 | 0.015 | 2.17 | 0.97 to 4.84 | 0.059 |
| Hodgkin's lymphoma | 0.78 | 0.19 to 3.22 | 0.74 | 1.17 | 0.28 to 4.96 | 0.83 |
| Leukaemia | 0.64 | 0.26 to 1.61 | 0.34 | 0.78 | 0.29 to 2.11 | 0.63 |
| Lung | 4.93 | 3.35 to 7.27 | <0.001 | 4.58 | 2.56 to 8.21 | <0.001 |
| Melanoma | 2.60 | 0.94 to 7.22 | 0.066 | 1.97 | 0.68 to 5.68 | 0.21 |
| Mesothelioma | 3.82 | 1.52 to 9.61 | <0.001 | 3.81 | 1.33 to 10.87 | 0.013 |
| Myeloma | 2.91 | 1.63 to 5.18 | <0.001 | 2.61 | 1.35 to 5.03 | 0.0043 |
| Non-Hodgkin's lymphoma | 1.93 | 1.19 to 3.14 | 0.0082 | 1.80 | 1.05 to 3.10 | 0.033 |
| Ovarian | 3.81 | 2.40 to 6.05 | <0.001 | 4.12 | 2.20 to 7.72 | <0.001 |
| Pancreatic | 6.23 | 3.61 to 10.74 | <0.001 | 4.57 | 2.44 to 8.55 | <0.001 |
| Prostate | 2.75 | 1.49 to 5.08 | <0.001 | 1.60 | 0.78 to 3.31 | 0.20 |
| Testicular | 0.000024 | 2.73E−102 to 2.17E+92 | 0.93 | 0.000042 | 1.26E−103 to 1.37E+94 | 0.93 |
| Endometrial | 2.52 | 1.37 to 4.64 | 0.0030 | 2.56 | 1.25 to 5.22 | 0.010 |

VTE, venous thromboembolism.

the tumour was available in the pathology database), and 50% were given radiotherapy (table 3). Breast cancer (n=2277) was the most common diagnosis followed by lung cancer (n=981) and colorectal cancer (n=567). The different clinical variables (gender, age, surgery, radiotherapy, platinum-based chemotherapy, distant metastases) divided according to cancer types are described in detail in table 3. About one-third (32.2%) of patients were treated with platinum-based chemotherapy during the disease course.

The risk of VTE increased only slightly with age both in univariate (HR 1.03, 95% CI 1.02 to 1.04, p<0.001) and multivariate analyses (HR 1.02, 95% CI 1.01 to 1.03, p=0.001). Patients who had disseminated cancer (distant metastases according to the ICD-10 coding) had an increased risk of VTE (HR 1.87, 95% CI 1.4 to 2.5, p<0.001) (table 4), as compared with all patients with cancer treated with chemotherapy.

Platinum-based chemotherapy was associated with increased VTE risk (HR 1.77, 95% CI 1.40 to 2.24, p<0.001), but after adjustment for other variables in the multivariate analysis, the association became insignificant (HR 0.83, 95% CI 0.56 to 1.21, p=0.33). The risk of VTE varied among the different cancer types, where the highest risks associated with lung cancer (HR 4.58, 95% CI 2.56 to 8.21, p<0.001) and pancreatic cancer (HR 4.57, 95% CI 2.44 to 8.55, p<0.001), followed by ovarian cancer (HR 4.12, 95% CI 2.20 to 7.72, p<0.001) and mesothelioma (HR 3.81, 95% CI 1.33 to 10.87, p=0.013) in multivariate analysis, as compared with breast cancer. Breast cancer was selected as the reference group due to the low rates of VTE events (n=206) among all chemotherapy-treated patients with breast cancer (n=7132) (table 2).

The estimated rate of VTE events in breast cancer based on Kaplan-Meier analysis was 12 per 1000 patients per year

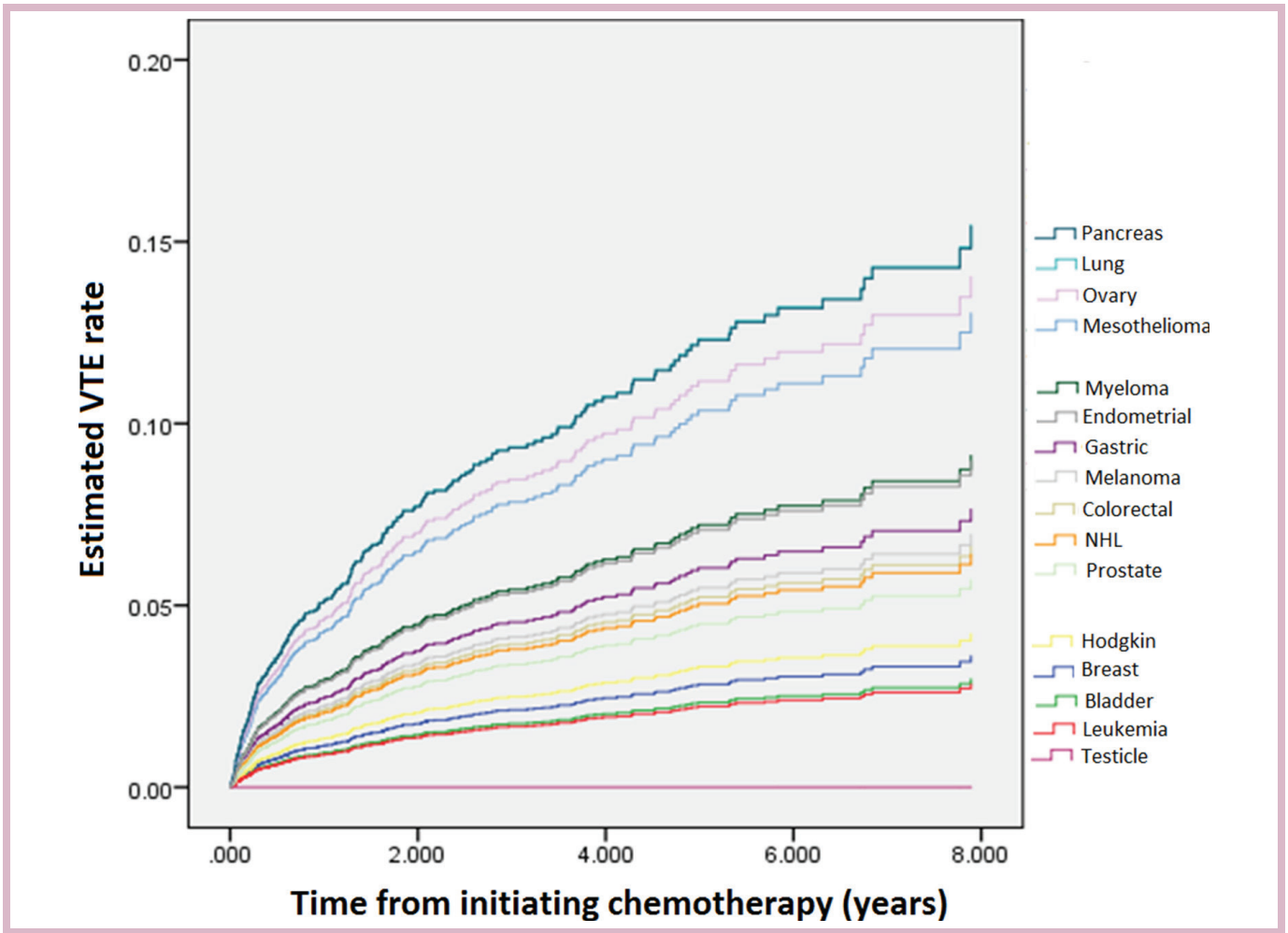


**Figure 2** Kaplan-Meier analysis of venous thromboembolism (VTE)-free time in chemotherapy-treated patients with cancer (n=7132) after the onset of chemotherapy. NHL, non-Hodgkin's lymphoma.

(figure 2). Within chemotherapy-treated patients with cancer, the rates of VTE events also varied in different cancer types. The risk of VTE can be grouped into three different types: high risk of VTE was documented in lung, pancreatic and ovarian cancers and in mesothelioma, while medium risk of VTE was observed in myeloma, endometrial, gastric, melanoma, colorectal, non-Hodgkin's lymphoma and prostate cancers. Hodgkin's lymphoma, leukaemia, breast, bladder and testicular cancers constituted a group of low risk of VTE in this study (figure 2). The risk of VTE was continued to increase during the entire course of disease and was not limited only to beginning of chemotherapy.

### Laboratory values and VTE risk

Based on ROC analysis, abnormal values of four laboratory variables were indicative of a VTE event: B-PLT (>316×$10^9$/L), total B-Leuk (>6.3×$10^9$ cells/L), B-Neut (>3.3×$10^9$ cells/L) and P-Crea (>62.5 µmol/L). Other laboratory variables, for example, B-Hgb level, E-MCV and E-MCH, and plasma electrolyte levels (P-Na and P-K), did not differ in patients with cancer with or without VTE.

Elevated B-Leuk (>6.3×$10^9$ cells/L) associated with VTE in the univariate analysis (HR 1.84, 95% CI 1.43 to 2.38, p<0.001), but in the multivariate analysis adjusted for additional laboratory variables (B-PLT, B-Neut and P-Crea) lost its significance (HR 1.38, 95% CI 0.97 to 1.95, p=0.07) (table 5). Similarly, elevated B-PLT (>316×$10^9$/L) associated in the univariate analysis with VTE (HR 1.57, 95% CI 1.23 to 1.99, p<0.001), but not in the multivariate analysis. In contrast, elevated B-Neut (>3.3×$10^9$ cells/L) continued to associate with VTE both in the univariate (HR 2.53, 95% CI 1.86 to 3.44, p<0.001) and the multivariate analysis (HR 1.96, 95% CI 1.33 to 2.89, p<0.001). Similarly, P-Crea level (>62.5 µmol/L) associated with VTE both in the univariate (HR 1.52, 95% CI 1.16 to 1.99, p=0.0021) and the multivariate analysis (HR 1.60, 95% CI 1.21 to 2.13, p=0.001).

## DISCUSSION

Combination of comprehensive data sets from various longitudinal electronic health records provides a powerful tool for identification of clinically meaningful associations that may have remained unnoticed. Here, we

Table 5 HRs of the association of VTE and the four major laboratory values by univariate and multivariate Cox regression analysis of the patients treated with chemotherapy (n=7778)

| | HR | Univariate 95% CI | P values | HR | Multivariate 95% CI | P values |
|---|---|---|---|---|---|---|
| Leucocyte count (over $6.3\times10^9$ cells/L) | 1.84 | 1.43 to 2.38 | <0.001 | 1.38 | 0.97 to 1.95 | 0.070 |
| Neutrophil count (over $3.3\times10^9$ cells/L) | 2.53 | 1.86 to 3.44 | <0.001 | 1.96 | 1.33 to 2.89 | <0.001 |
| Platelet count (over $316\times10^9$ cells/L) | 1.57 | 1.23 to 2.00 | <0.001 | 1.22 | 0.93 to 1.59 | 0.15 |
| Plasma creatinine (over 62.5 µmol/L) | 1.52 | 1.16 to 1.99 | 0.0021 | 1.60 | 1.21 to 2.13 | 0.0010 |

VTE, venous thromboembolism.

used an extensive data set of almost 500 000 hospitalised patients covering 9-year follow-up during years 2005–2013 to characterise comorbidities associated with VTE events, with special emphasis on the relationship between cancer and VTE. The individual records included ICD-10 codes, SNOMED, clinical laboratory results and structured treatment information. With this real-world information, we evaluated the incidence of VTE events within the hospital patients (both inpatient and outpatient settings) and the specific features among patients with cancer with special emphasis on hypothesis-free associations rather than established risk factors. The 'big data' approach from medical records appears well suited for identification of clinically meaningful associations, such as comorbidities. While our approach cannot be used to demonstrate direct causalities, we find expected parallels, including increased VTE incidence among patients with cardiovascular disease (including atrial fibrillation/flutter, coronary heart disease, hypertension, cerebrovascular disease or congestive heart failure) and obesity. Interestingly, increased VTE incidence among patients with psychiatric disorders was noticed. Antipsychotic drugs have been associated with increased risk of VTE,[27] which may, at least partially, contribute to our findings. Among major ICD disease categories, patients with cancer had an increased VTE risk, and in several cancer types and in progressed disease states, more than 5% of patients experienced a VTE event. Our relatively low VTE rate as compared with other reported studies[22] might be explained by cohort differences. For instance, our study included a large set of cancer types, and both patients who received or did not receive chemotherapy were included. Also, the lack of information on thromboprophylaxis hampers the VTE incidence comparisons difficult.

In further analyses of the cancer cohort, platinum-based chemotherapy was a risk factor for VTE in univariate analysis, but not in multivariate analysis. The predictive value of single variable can be lost after adjustment for established variables in a multivariable model for the possible relationship between other variables and VTE. In platinum-based chemotherapy the VTE risk is not class specific for all the platinum-containing chemotherapy agents, which may influence estimation of VTE risk confounded by different cancer types. Especially cisplatin-based chemotherapy is associated with increased risk of VTE.[28 29] Previously, Starling *et al* reported a difference in the incidence of thromboembolic events in the cisplatin-containing regimens, as compared with the oxaliplatin-containing regimens in patients with advanced gastro-oesophageal cancer.[30]

VTE in patients with cancer is associated with increased mortality[31] and for the clinical impact, VTE complications in patients with cancer cause significant economic burden.[32] It is therefore of both individual and economic interest to identify patients with cancer with the highest VTE risk, who could benefit from thromboprophylaxis. Routine prevention cannot be recommended for all patients with cancer, as anticoagulant treatment enhances bleeding tendency,[33] but it should be considered for carefully selected high-risk patients.[34–36] Different biomarkers, for example, soluble P-selectin and D-dimer, have been specifically investigated for their capacity of predicting VTE during the course of disease,[37] but these markers were unfortunately not routinely measured during the study period.

Also, laboratory variables such as blood cell counts (elevated B-Leuk and B-PLT and decreased B-Hgb) have been evaluated to be useful in risk prediction during chemotherapy by the Khorana score.[38] We evaluated the association of available laboratory values in chemotherapy-treated patients with VTE events, and found that B-Leuk, B-PLT, B-Neut and creatinine levels are associated with VTE risk. Khorana score is one of the most established models for predicting VTE risk during chemotherapy. The Khorana score includes three laboratory values, B-PLT ($>350\times10^9$/L), B-Hgb level (<100 g/L), B-Leuk ($>11\times10^9$/L), which indicate the probability of a VTE event associated with chemotherapy. In addition, the cancer type and body mass index, which was not available in our database, are also included in Khorana score.[38] Apart from Khorana score, we also evaluated the association of 10 other laboratory values with VTE, including E-MCV and E-MCH. Our retrospective study results differ from Khorana regarding the predictive value of the elevated B-Leuk ($>6.3\times10^9$ cells/L), elevated B-PLT ($>316\times10^9$/L), elevated B-Neut ($>3.3\times10^9$ cells/L)

and P-Crea level (>62.5 µmol/L) in patients with cancer treated with chemotherapy. These laboratory variables could have implications for thromboprophylaxis in patients with high risk of VTE during chemotherapy. Unfortunately, we did not have functional renal assessment available. Interestingly, reduced estimated glomerular filtration rate even under normal serum creatinine values has been shown to associate with an increased risk of VTE.[39]

To the best of our knowledge, this is the first cohort study that combines a wide range of different longitudinal hospital electronic medical records to provide real-world analyses of VTE risk factors. Corraini *et al* used Danish nationwide medical database to survey the impact of comorbidity on VTE risk, but they concentred only on patients with stroke.[40] Our large retrospective study has some limitations, including lack of information on some previously known VTE risk factors. Compatible with the real-world setting, the clinical practice of diagnosing and especially documenting VTE events is likely to be suboptimal. We did not have sufficient data on anticoagulation therapy, which might have been useful in validating the VTE diagnosis. However, the size of the data set and the correlation of overall results with previous publications suggest that our approach is valid. We lacked genetic data on inherited and acquired thrombophilias (eg, factor V Leiden mutation), which have been documented to increase the risk of VTE also in patients with cancer.[41] As the known inherited conditions are uncommon (5% in our population) their influence on our result remains only marginal.

Hospital registries offer a targeted opportunity to study risk of VTE in large epidemiological cohort of patients with cancer. This approach uses baseline clinical and laboratory variables, and in the biobank setting, can also link electronic medical information to biological samples, at minimum using specimens collected for diagnostic purposes. In conclusion, our study demonstrates that compilation of large data sets combining multiple longitudinal electronic medical databases for research purposes is feasible and provides useful real-world data and novel tools for discoveries. It also provides tools for further dissecting both clinical and biological factors associated with cancer-related VTE events and aids at targeting thromboprophylaxis.

**Contributors** MHP collected and interpreted the data and wrote the manuscript with OMC. SK interpreted the data, performed the statistical analyses and drafted the figures. RL interpreted the data and provided clinical expertise. OMC interpreted the data and wrote the manuscript with MHP. All authors were involved in study concept and design, contributed to the manuscript and approved the final version.

**Funding** This work was supported by grants from the Finnish Cancer Society and Turku University Hospital Research Funds.

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** The Turku Clinical Research Center (permission number T133/2013).

**Provenance and peer review** Not commissioned; externally peer reviewed.



## REFERENCES

1. Kearon C. Natural History of Venous Thromboembolism. *Circulation* 2003;107(90231):22–.
2. 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 Intern Med* 1998;158:585–93.
3. Smeeth L, Cook C, Thomas S, *et al*. Risk of deep vein thrombosis and pulmonary embolism after acute infection in a community setting. *Lancet* 2006;367:1075–9.
4. Heit JA, Silverstein MD, Mohr DN, *et al*. Risk factors for deep vein thrombosis and pulmonary embolism: a population-based case-control study. *Arch Intern Med* 2000;160:809–15.
5. Schulman S, Lindmarker P, Holmström M, *et al*. Post-thrombotic syndrome, recurrence, and death 10 years after the first episode of venous thromboembolism treated with warfarin for 6 weeks or 6 months. *J Thromb Haemost* 2006;4:734–42.
6. Spencer FA, Gore JM, Lessard D, *et al*. Patient outcomes after deep vein thrombosis and pulmonary embolism: the Worcester Venous Thromboembolism Study. *Arch Intern Med* 2008;168:425–30.
7. Geerts WH, Bergqvist D, Pineo GF, *et al*. *Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines*. 8th Edition: Chest, 2008:133;453S.
8. Kahn SR, Morrison DR, Cohen JM, *et al*. Interventions for implementation of thromboprophylaxis in hospitalized medical and surgical patients at risk for venous thromboembolism. *Cochrane Database Syst Rev* 2013;77:CD008201.
9. Wells PS, Forgie MA, Rodger MA. Treatment of venous thromboembolism. *JAMA* 2014;311:717–28.
10. Sallah S, Wan JY, Nguyen NP. Venous thrombosis in patients with solid tumors: determination of frequency and characteristics. *Thromb Haemost* 2002;87:575–9.
11. Chew HK, Wun T, Harvey D, *et al*. Incidence of venous thromboembolism and its effect on survival among patients with common cancers. *Arch Intern Med* 2006;166:458–64.
12. Rickles FR, Falanga A. Activation of clotting factors in cancer. *Cancer Treat Res* 2009;148:31–41.
13. Timp JF, Braekkan SK, Versteeg HH, *et al*. Epidemiology of cancer-associated venous thrombosis. *Blood* 2013;122:1712–23.
14. Wun T, White RH. Epidemiology of cancer-related venous thromboembolism. *Best Pract Res Clin Haematol* 2009;22:9–23.
15. Blom JW, Vanderschoot JP, Oostindiër MJ, *et al*. Incidence of venous thrombosis in a large cohort of 66,329 cancer patients: results of a record linkage study. *J Thromb Haemost* 2006;4:529–35.
16. Haddad TC, Greeno EW. Chemotherapy-induced thrombosis. *Thromb Res* 2006;118:555–68.
17. Khorana AA, Francis CW, Culakova E, *et al*. Risk factors for chemotherapy-associated venous thromboembolism in a prospective observational study. *Cancer* 2005;104:2822–9.
18. Sørensen HT, Mellemkjaer L, Olsen JH, *et al*. Prognosis of cancers associated with venous thromboembolism. *N Engl J Med* 2000;343:1846–50.
19. Dickmann B, Ahlbrecht J, Ay C, *et al*. Regional lymph node metastases are a strong risk factor for venous thromboembolism: results from the Vienna Cancer and Thrombosis Study. *Haematologica* 2013;98:1309–14.
20. Ording AG, Nielsen ME, Smith AB, *et al*. Venous thromboembolism and effect of comorbidity in bladder cancer: A danish nationwide cohort study of 13,809 patients diagnosed between 1995 and 2011. *Urol Oncol* 2016;34–292.e1–292.e8.
21. Smith AB, Horvath-Puhó E, Nielsen ME, *et al*. Effect of comorbidity on risk of venous thromboembolism in patients with renal cell carcinoma. *Urol Oncol* 2014;32:466–72.
22. Khorana AA, Dalal M, Lin J, *et al*. Incidence and predictors of venous thromboembolism (VTE) among ambulatory high-risk cancer patients undergoing chemotherapy in the United States. *Cancer* 2013;119:648–55.
23. Ratasvuori M, Lassila R, Laitinen M. Venous thromboembolism after surgical treatment of non-spinal skeletal metastases - An underdiagnosed complication. *Thromb Res* 2016;141:124–8.

24. Kaatz S, Fu AC, AbuDagga A, *et al*. Association between anticoagulant treatment duration and risk of venous thromboembolism recurrence and bleeding in clinical practice. *Thromb Res* 2014;134:807–13.
25. Ferroni P, Zanzotto FM, Scarpato N, *et al*. Risk Assessment for Venous Thromboembolism in Chemotherapy-Treated Ambulatory Cancer Patients. *Med Decis Making* 2017;37:234–42.
26. Ferroni P, Zanzotto FM, Scarpato N, *et al*. Validation of a Machine Learning Approach for Venous Thromboembolism Risk Prediction in Oncology. *Dis Markers* 2017;2017:1–7.
27. Lacut K, Le Gal G, Couturaud F, *et al*. Association between antipsychotic drugs, antidepressant drugs and venous thromboembolism: results from the EDITH case-control study. *Fundam Clin Pharmacol* 2007;21:643–50.
28. Seng S, Liu Z, Chiu SK, *et al*. Risk of venous thromboembolism in patients with cancer treated with Cisplatin: a systematic review and meta-analysis. *J Clin Oncol* 2012;30:4416–26.
29. Moore RA, Adel N, Riedel E, *et al*. High incidence of thromboembolic events in patients treated with cisplatin-based chemotherapy: a large retrospective analysis. *J Clin Oncol* 2011;29:3466–73.
30. Starling N, Rao S, Cunningham D, *et al*. Thromboembolism in patients with advanced gastroesophageal cancer treated with anthracycline, platinum, and fluoropyrimidine combination chemotherapy: a report from the UK National Cancer Research Institute Upper Gastrointestinal Clinical Studies Group. *J Clin Oncol* 2009;27:3786–93.
31. Khorana AA, Francis CW, Culakova E, *et al*. Thromboembolism is a leading cause of death in cancer patients receiving outpatient chemotherapy. *J Thromb Haemost* 2007;5:632–4.
32. Kourlaba G, Relakis J, Mylonas C, *et al*. The humanistic and economic burden of venous thromboembolism in cancer patients: a systematic review. *Blood Coagul Fibrinolysis* 2015;26:13–31.
33. Prandoni P, Trujillo-Santos J, Surico T, *et al*. Recurrent thromboembolism and major bleeding during oral anticoagulant therapy in patients with solid cancer: findings from the RIETE registry. *Haematologica* 2008;93:1432–4.
34. Watson HG, Keeling DM, Laffan M, *et al*. British Committee for Standards in Haematology. Guideline on aspects of cancer-related venous thrombosis. *Br J Haematol* 2015;170:640–8.
35. Lyman GH, Bohlke K, Khorana AA, *et al*. Venous thromboembolism prophylaxis and treatment in patients with cancer: american society of clinical oncology clinical practice guideline update 2014. *J Clin Oncol* 2015;33:654–6.
36. Farge D, Debourdeau P, Beckers M, *et al*. International clinical practice guidelines for the treatment and prophylaxis of venous thromboembolism in patients with cancer. *J Thromb Haemost* 2013;11:56–70.
37. Ay C, Dunkler D, Marosi C, *et al*. Prediction of venous thromboembolism in cancer patients. *Blood* 2010;116:5377–82.
38. Khorana AA, Kuderer NM, Culakova E, *et al*. Development and validation of a predictive model for chemotherapy-associated thrombosis. *Blood* 2008;111:4902–7.
39. Ferroni P, Guadagni F, Laudisi A, *et al*. Estimated glomerular filtration rate is an easy predictor of venous thromboembolism in cancer patients undergoing platinum-based chemotherapy. *Oncologist* 2014;19:562–7.
40. Corraini P, Ording AG, Henderson VW, *et al*. Cancer, other comorbidity, and risk of venous thromboembolism after stroke: a population-based cohort study. *Thromb Res* 2016;147:88–93.
41. Blom JW, Doggen CJ, Osanto S, *et al*. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. *JAMA* 2005;293:715–22.