SOHO ANNUAL MEETING 2024

ANALYSIS OF CARDIOVASCULAR RISK IN 920 PATIENTS WITH MYELOPROLIFERATIVE NEOPLASMS USING NATURAL LANGUAGE PROCESSING

A. DUMINUCO[1,2], J. AU YEUNG[1], S. VIRDEE[1], R. VAGHELA[1], C. WOODLEY[1], S. ASIRVATHAM[1], N. CURTO-GARCIA[1], P. SRISKANDARAJAH[1], J. O'SULLIVAN[1], D. RADIA[1], H. DE LAVALLADE[1], S. KORDASTI[1], G. PALUMBO[2], C. HARRISON[1] and P. HARRINGTON[1]

1. Guys and St Thomas NHS Foundation Trust, London, United Kingdom
2. AOU Policlinico "G.Rodolico-San Marco", Catania, Italy

NHS Guy's and St Thomas' NHS Foundation Trust

INTRODUCTION

Thromboembolic events represent the most common cause of morbidity and mortality in myeloproliferative neoplasms (MPN). Natural language processing is a branch of machine learning involving computational interpretation and human language analysis. CogStack is an information extraction architecture incorporating structured and unstructured electronic health record (EHR) components.

AIM

To employ a machine-learning approach to determine the prevalence of cardiovascular risk factors and association with development of thrombotic events.

METHODS

Data extracted from CogStack was processed by a medical concept annotation toolkit (MedCAT).
Using deep learning, MedCAT was used to disambiguate and capture synonyms and acronyms for Systematized Nomenclature of Medicine–Clinical Terms concepts, whilst also assessing linguistic context.
The base MedCAT model was trained in an unsupervised manner on >18 million EHR documents and this was further fine-tuned through 500 clinician annotated MPN specific documents. MedCAT mapped mentions of relevant concepts to respective SNOMED-CT codes and total counts were aggregated and grouped by individual patient.

RESULTS

Data from 360 polycythaemia vera (PV) and 560 essential thrombocythaemia (ET) patients seen at Guys' and St Thomas NHS Foundation Trust (GSTT) between 2005 and April 2023 were evaluated, including 24,155 individual EHR documents.

- In ET, hypertension (HTN) was the most prevalent comorbidity, identified in 21.3% (119) of patients, followed by hypercholesterolemia (9.6%) and smoking (6.6%), while, overall, 20% (112) experienced a thrombotic event.
- This included thrombosis not otherwise specified (NOS) in 8% (45), cerebrovascular accident (CVA) in 7.7% (43), and myocardial infarction (MI) in 3.6% (20).
- HTN was also the most identified condition in PV, seen in 23.1% (83) of patients. In PV, thrombosis NOS was observed in 19.4% (70), CVA in 14.2% (51), venous thromboembolism (VTE) in 23.3% (84) and MI in 3.1% (11).
- Overall, 35% (126) of cases had a thrombotic event. Comparing the two cohorts, a significantly greater frequency of events was observed in PV patients for CVA (p=0.002), portal vein thrombosis, and VTE (both p<0.001) compared to ET (Table 1).

Table 1. Occurrence of any type of thrombotic event in ET and PV patients. NOS: not otherwise specified

Type of thrombosis	ET (560 patients) n (%)	PV (360 patients) n (%)	*p*
Deep venous thrombosis	8 (1.4)	10 (2.8)	0.15
Pulmunary embolism	10 (1.8)	10 (2.8)	0.314
Myocardial infarction	20 (3.6)	11 (3.1)	0.673
Cerebrovascular accident	43 (7.7)	51 (14.2)	*0.002*
Portal vein thrombosis	7 (1.3)	18 (5)	*<0.001*
Cerebral venous thrombosis	6 (1.1)	1 (0.3)	0.177
Thrombosis, NOS	45 (8)	70 (19.4)	*<0.001*
Venous thromboembolism	65 (11.6)	84 (23.3)	*<0.001*
Overall thrombotic events	112 (20)	126 (35)	*<0.001*

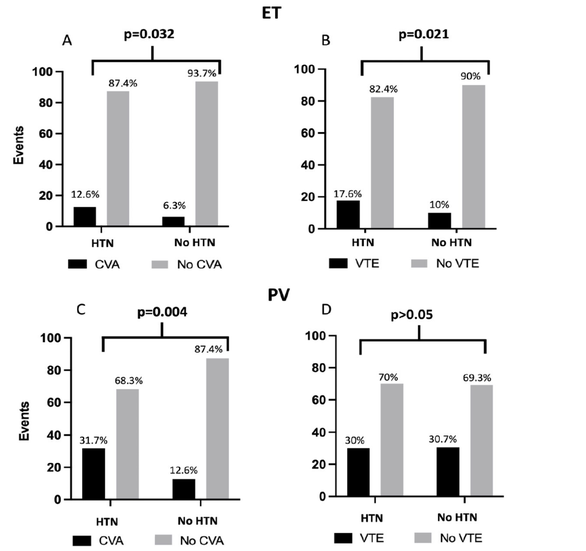


Figure 2. Relationship between cardiovascular risk factor and type of thrombotic event, with a statistically significant correlation between HTN and CVA, in both ET (A) and PV (C) cohorts. Increased incidence of VTE in HTN patients also observed in ET cohort (B)

- ET patients with HTN were more likely to have CVA (of 119 HTN affected, 15 had CVA, and of 441 HTN-negative, 28 had CVA, p=0.032) and VTE than those without (of 119 HTN-affected, 21 had VTE, and of 441 HTN-negative, 44 had VTE, p=0.021).
- Similarly, PV patients with HTN were also more likely to develop CVA (20:63 vs 31:246, p=0.004).
- On multivariate analysis, HTN was confirmed as having a central role in increasing the risk of any type of thrombotic event in both ET and PV cohorts (OR 2.5; 95% CI 1.5-4.2; p<0.001 in ET, and 1.5; 95%CI 1.1-2.8; p<0.016 in PV).

CONCLUSIONS

We describe a novel machine learning approach to assess cardiovascular comorbidities in patients with MPN, allowing big data analysis, with the potential to better define thrombotic risk.

We provide a rare 'real-world' report on the prevalence of comorbidities in this group, confirming increased CVA and VTE in patients with HTN, emphasising the importance of strict management of this comorbidity.

In keeping with previous smaller studies, we also confirm a significantly higher risk of thrombotic events in patients with a diagnosis of PV when compared with ET.

CONTACTS

Dr. Andrea Duminuco
Dr Patrick Harrington

andrea.duminuco@gstt.nhs.uk
patrick.harrington@gstt.nhs.uk