

# Identifying Risk Groups in 502 Patients with Total Hip Arthroplasty and Deep Vein Thrombosis: A Machine Learning Clustering Analysis

Alejandra Chavarria, Oserekpamen Favour Omobhude, Alishah Ahmadi, Anthony J Kaywood, Adam C Kiss, Mateusz Faltyn, Jason Shih Hoellwarth

## INTRODUCTION:

Deep vein thrombosis (DVT) is a serious complication that may occur after a total hip arthroplasty (THA) and can lead to life-threatening conditions such as pulmonary embolism if not promptly diagnosed and managed. Understanding the risk factors and patient characteristics associated with DVT is essential for improving surgical outcomes and postoperative care. This study uses machine learning based clustering algorithms to identify unique patient profiles, providing insight into how certain comorbidities affect outcomes in patients with an acquired DVT after undergoing THA. Identifying patterns within these patient subgroups may support early intervention efforts to mitigate the risk of adverse events in high-risk individuals.

## METHODS:

The 2015-2021 National Inpatient Sample was queried using ICD-10 CM/PCS codes to identify patients who had a THA and acquired a DVT during the same hospital stay. Machine-learning clustering analysis evaluated the cohort based on 49 comorbidities, complications, and clinical covariates to assess their association with a non-routine discharge (NRD). The optimal number of clusters was determined using the Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI). Between-cluster multivariate logistic regression analysis was performed to assess risk of non-routine discharge. Kruskal-Wallis H-Testing was performed to assess variance in length-of-stay between clusters. Statistical analysis was performed using Python, noting  $p < .05$  as significant.

## RESULTS:

502 patients were included. DBI-CHI composite determined the optimal number of clusters to be 9; however, clusters 8 and 9 were each comprised of less than 10 patients so they were not included in this study. Most clusters demonstrated a high prevalence of osteoarthritis, hyperlipidemia, and hypertension. Cluster 1 had the lowest rate of NRD (59.38%) and the lowest prevalence of comorbidities. Cluster 2 had a NRD of 69.78% and a high prevalence of hypothyroidism. Cluster 3 had a NRD of 76.0% and the highest prevalence of obesity. Cluster 4 had a NRD of 79.26% and a high prevalence of arrhythmias. Cluster 5 had a NRD of 89.47% and a high prevalence of anemia and diabetes. Cluster 6 had a NRD of 91.67% and a high prevalence of anxiety, arrhythmia, diabetes, and hypothyroidism. Cluster 7 had the highest rate of NRD (92.31%) and the highest prevalence of chronic kidney disease. The odds ratio of NRD for Cluster 7 was 8.21 ( $p < 0.05$ ) compared to Cluster 1.

## DISCUSSION AND CONCLUSION:

The clustering analysis used machine learning to identify 7 distinct patient profiles among individuals who developed DVT following THA. Cluster 1 exhibited the lowest NRD rate and the fewest comorbidities, representing a low risk group. In contrast, Cluster 7 had the highest NRD rate and the greatest burden of chronic kidney disease, highlighting a significantly high risk population. The association between increased comorbidity burden and NRD emphasizes the need for targeted protocols and interventions to reduce postoperative complications and improve overall patient outcomes.

