Causal Insights into Stroke Mortality Risk Reduction

Dalia Kriksciuniene, Virgilijus Sakalauskas
Kauno Kolegija, Lithuania

Abstract

Stroke is a disease with severe and lasting consequences for patients' health, which makes research into effective treatment methods essential for reducing its impact. In this study, we employ causal inference to gain a deeper understanding of the real effect of treatment methods on stroke mortality. We considered 8 variables from 944 stroke patients treated at the Clinical Centre of Montenegro, including treatment methods, age, health status, and stroke type. To reduce bias and create fair comparisons, we applied Propensity Score Matching (PSM), which allowed us to build groups of patients with similar baseline characteristics. To understand how treatments worked differently for different patients, we also used causal forests, an advanced ensemble method designed to discover variation in treatment effects across subgroups. Our analysis revealed that Age, Health Status, and Past Stroke were the most significant factors. Some interventions reduced the risk of death by up to 14% compared to alternatives. These findings demonstrate that causal inference provides more clinically actionable insights than predictive models alone.

Introduction

Understanding how different treatment strategies causally influence stroke-related mortality, rather than merely correlate with outcomes, is critical to improving patient survival and personalising care.

This distinction between correlation and causation is not trivial: interventions based on non-causal models risk reinforcing existing biases in care. Consequently, there is growing interest in applying causal inference methods for observational healthcare data to estimate treatment effects with greater validity.

Causal inference bridges the gap between prediction and explanation. Among these methods, Propensity Score Matching (PSM) has emerged as a foundational technique. Initially proposed by Rosenbaum and Rubin (1983), PSM addresses selection bias in observational studies by creating quasi-randomised groups with similar distributions of covariates. When appropriately applied, PSM enables researchers to compare treatment effects as if they originated from a randomised controlled trial, thereby approximating experimental conditions using real-world data. In stroke research, this is particularly useful where randomisation is ethically or logistically infeasible.

However, PSM by itself is limited to average treatment effects and does not accommodate the complexity of treatment effect heterogeneity. Introduced by Wager and Athey (2018), causal forests extend the random forest algorithm to estimate Conditional Average Treatment Effects (CATEs) at the individual level. This not only quantifies the expected benefit of a treatment for a given patient profile but also helps to identify which subgroups benefit most—an essential step toward personalised medicine.

Our work contributes to this emerging field by applying a combined PSM and causal forest approach to real-world stroke data. This combination of balancing groups through PSM and capturing nuanced effects through causal forests provides a more refined and clinically relevant understanding of how different interventions influence stroke-related mortality. Unlike global models that output a single effect size, causal forests highlight variability in treatment responses across patients, offering clinicians not just a better average, but a roadmap for targeted decision-making.

Stroke treatment strategies

A stroke occurs when the blood supply to part of the brain is disrupted, leading to potential brain damage. Broadly, strokes are categorized into two main types: ischemic and hemorrhagic. Ischemic strokes, which account for approximately 87% of all cases, result from a blockage in the blood vessels supplying the brain. Hemorrhagic strokes involve bleeding in or around the brain and are divided into two categories based on the location and source of bleeding:

- Intracerebral hemorrhage occurs when a blood vessel ruptures within the brain tissue itself.
- Subarachnoid hemorrhage (SAH) involves bleeding in the space between the brain and the surrounding membranes, typically caused by a ruptured aneurysm.

SAH represents about 5–6% of all stroke cases and is classified separately due to its distinct origin and clinical features. Other stroke types include cryptogenic stroke, brain stem stroke, and various less frequent forms.

In this study, we compare the effectiveness of two stroke treatment strategies:

- ANTIKOAG (Anticoagulants): These drugs target clotting factors to prevent the formation of blood clots. Examples include warfarin, heparin, and DOACs. They are typically used for conditions such as atrial fibrillation or venous thromboembolism.
- ANTIAGREGAC (Antiplatelets): These agents, such as aspirin and clopidogrel, inhibit platelet aggregation and are commonly prescribed after ischemic strokes or myocardial infarction.

Both therapies aim to reduce the risk of stroke but operate at different stages of the clotting cascade and are used in distinct clinical contexts.

Our goal for this research is to compare these two treatment types—coded as 1 (anticoagulants) and 2 (antiplatelets) - against each other, as well as against other treatment methods not falling into either category, which are coded as 0.

Experimental Data structure

The database applied for evaluating treatment effects on stroke-related survival consists of stroke patients records from neurology department of the Clinical Centre of Podgorica, Montenegro collected between February 25, 2017, and December 18, 2019. After cleansing and selecting adequate variables from the original database we selected structured 944 records of stroke patients and 8 variables. The data of stroke patients varies by age (13 to 96 years) and gender (485-male, 427-female). The short example of database structure and data records is presented in Table below.

Variable name	Meaning and coding of data
Vital Status	– 0:Event (death), 1: Alive
Stroke Type	- 1: Ischemic, 2: Hemorag, 3: SAH, 4: Unspecified
Treatment methods	- 0:other treatments,1:Anticoagulation, 2:Antiplatelet Therapy
Health Status	- Health score before stroke from 0:best to 9:worst: 0: Without symptoms; 1: Without significant disability despite symptoms; 2: Minor disability; 3: Moderate disability, but able to walk independency; 4: Moderate disability, not able to walk independency; 5: Major disability; 9: Unknown
Age	– Patient age, years
Gender	– 1:Male, 2:Female, 9:Unspecified
Past Stroke	– Stroke in the past. 1:Yes, registered in the patient health record, 0:No
Smoke Status	1-Smokes, 2-No, 3-Smoked before

Methods description

Propensity Score Matching (PSM)

A propensity score represents the probability that a patient would receive a specific treatment, based on characteristics such as age, gender, stroke type, past health conditions, and smoking status. We calculated these scores using logistic regression, then matched patients who received one treatment with similar patients who received another. To make sure the matches were good, we set a limit (called a caliper) so that patients were only paired if their scores were close enough.

After matching, the groups were more balanced. That means their baseline characteristics looked very similar, which helps us focus on the effect of the treatment itself - not the differences between patients.

Causal Forests

Once we had balanced treatment groups, we wanted to go further and understand how treatment effects might differ across individual patients. For this, we used a method called Causal Forests, which is based on decision trees.

Causal Forests work by splitting the data into many small groups where patients are similar, and then estimating the treatment effect within each group. These small trees are combined into a large "forest" that provides an estimate of how much a treatment helps or harms each patient- this is known as the Conditional Average Treatment Effect (CATE).

Treatments and Python libraries

We evaluated the causal effect of three treatment strategies on stroke-related mortality using matched datasets and Causal Forest analysis. The treatments compared were:

- Treatment 2: Antiplatelet therapy (ANTIAGREGAC)
- Treatment 1: Anticoagulant therapy (ANTIKOAG)
- Treatment 0: Other medications or no targeted treatment

Each comparison included a Propensity Score Matching (PSM) step to ensure balanced cohorts, followed by estimation of individual treatment effects using Causal Forests.

All analyses from our data were conducted using Python libraries. For implementing Propensity Score Matching (PSM), we used pandas for data loading and manipulation, numpy for numerical operations, and scikit-learn (sklearn) to build the logistic regression model used to estimate propensity scores. Visualization of covariate balance before and after matching was performed using matplotlib and seaborn.

To estimate heterogeneous treatment effects, we applied the Causal Forest method using the CausalForestDML class from the econml library—one of the most established implementations of causal forests in Python. The scikit-learn library was also used within this framework for modeling and preprocessing. Matplotlib and seaborn were again employed to visualize CATE distributions, scatterplots, and feature importance rankings.

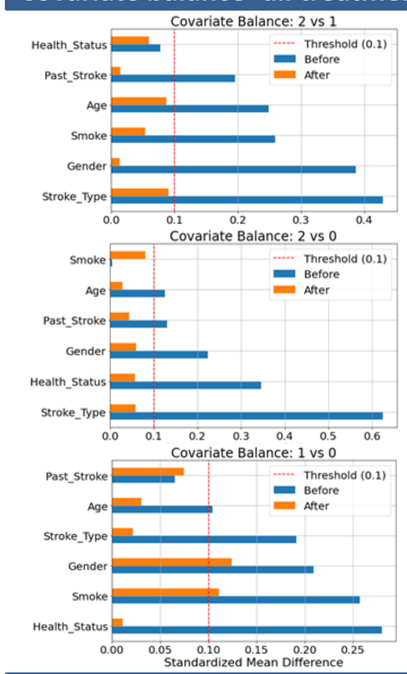
Results of calculations

Below we present the results of our calculations. The table shows number of patients after PSM and Average Treatment Effect (ATE). The number of matched pairs used for each analysis means the patients with similar characteristics from both treatment groups, selected using Propensity Score Matching to reduce bias. ATE here stands for the average difference in survival probability caused by the treatment. A positive ATE means the treatment increased survival; a negative ATE suggests the opposite.

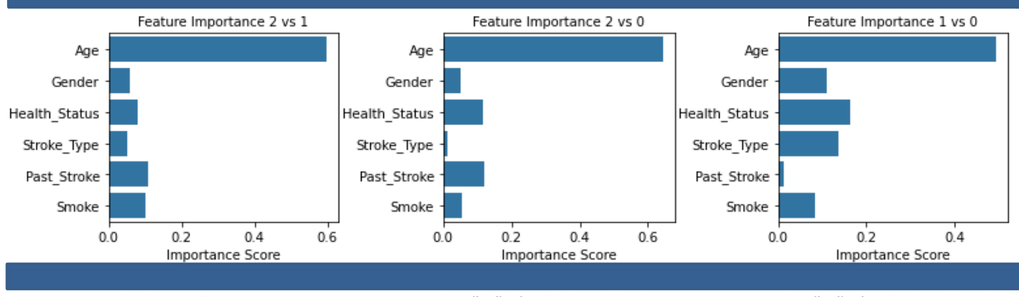
Comparison	Treatment Codes	Patients after PSM	ATE (Average Treatment Effect)
Antiplatelet vs Anticoagulant	2 vs 1	671	0.08597
Antiplatelet vs No Targeted Treatment	2 vs 0	675	0.15038
Anticoagulant vs No Targeted Treatment	1 vs 0	219	–0.06503

The *Antiplatelet therapy* (treatment 2) consistently showed the most positive and reliable impact on survival after stroke, both compared to *Anticoagulants* and to *No specific treatment*. An approximately 8.6% survival chance is higher compared to *Anticoagulant* therapy (treatment 1), and 15% higher than *No targeted treatment* (treat. 0). On the other hand, *Anticoagulant* therapy showed no consistent benefit, with ATE at -6.5%, possibly resulting in some harm, depending on the patient's profile.

Covariate balance- all treatments

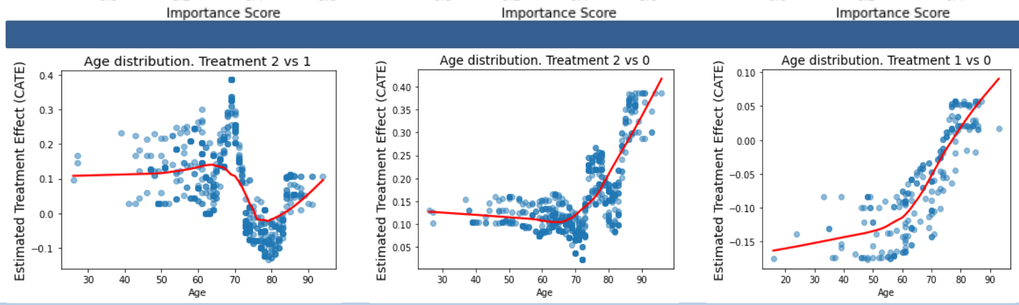


Graphical results interpretation



In Causal Forests, feature importance doesn't show which variables predict outcomes (like in standard models) - it indicates which variables help explain differences in treatment effects between patients.

Our model for all treatments ranked features in a very similar way. The most important are *Age, Health Status* and *Past Stroke.* Factors less influencing the differences in treatment effects are *Gender, Stroke Type* and *Smoking*.



Age is more than 3 times more important than any other feature. Each plot illustrates the relationship between age and the effectiveness of stroke treatments. The vertical axis shows the estimated benefit of treatment (CATE), and the horizontal axis is Age. Higher values mean more benefit. From 3 graph we see that Anticoagulants should not be used for patients younger than 80 years of age, as they reduce the likelihood of survival.

Conclusion remarks

This study employed Propensity Score Matching and Causal Forest models to assess the impact of various stroke treatment strategies on patient survival. The results show that Antiplatelet therapy consistently provided the greatest survival benefit, particularly among older patients. In contrast, Anticoagulant therapy showed limited or even negative effects, especially in younger individuals. These findings highlight the importance of personalised treatment decisions, as patient characteristics - especially Age - strongly influence treatment response. Causal forest analysis proved valuable in uncovering these individual-level differences, supporting more targeted and effective stroke care strategies.

Contacts dalia.kriksciuniene@go.kauko.lt
virgilijus.sakalauskas@go.kauko.lt

Kauno Kolegija
Faculty of Computing, Engineering and Technologies